

A second order time discretization of the solution of the non-linear filtering problem

Dan Crisan^{*}, Salvador Ortiz-Latorre[†]

*Department of Mathematics
Imperial College London
Huxley's Building
180 Queen's Gate
London SW7 2AZ
United Kingdom*

e-mail: d.crisan@imperial.ac.uk

*Centre of Mathematics for Applications
Oslo University
P.O. Box 1053 Blindern
N-0316 Oslo
Norway*

e-mail: s.ortiz-latorre@cma.uio.com

Abstract: The solution of the continuous time filtering problem can be represented as a ratio of two expectations of certain functionals of the signal process that are parametrized by the observation path. We introduce a new time discretisation of these functionals corresponding to a chosen partition of the time interval and show that the convergence rate of discretisation is proportional with the square of the mesh of the partition.

Primary 60F05, 60F25, 60G35, 60H35, 93E11.

Keywords and phrases: Nonlinear filtering, second order time discretisations, Kallianpur-Striebel's formula.

1. Introduction

Partially observed dynamical systems are ubiquitous in a multitude of real-life phenomena. The dynamical system is typically modelled by a continuous time stochastic process called the signal process X . The signal process cannot be measured directly, but only via a related process Y , called the observation process. The filtering problem is that of estimating the current state of the dynamical system at the current time given the observation data accumulated up to that time. Mathematically the problem entails computing the conditional distribution of the signal process X_t , denoted by π_t , given \mathcal{Y}_t , the σ -algebra generated by Y . In a few special cases, π_t can be expressed in closed form

^{*}The work of D. C. was partially supported by the EPSRC Grant EP/H0005500/1.

[†]The work of S. O-L. was supported by the BP-DGR 2009 grant and the project Energy Markets: Modeling, Optimization and Simulation (EMMOS), funded by the Norwegian Research Council under grant Evita/205328.

as a functional of the observation path. For example, the celebrated Kalman-Bucy filter does this in the linear case. In general, an explicit formula for π_t is not available and inferences can only be made by numerical approximations of π_t . As expected the problem has attracted a lot of attention in the last fifty years (see Chapter 8 of [1] for a survey of existing numerical methods for approximating π_t). Particle methods¹ are algorithms which approximate π_t with discrete random measures of the form $\sum_i a_i(t)\delta_{v_i(t)}$, in other words with empirical distributions associated with sets of randomly located particles of stochastic masses $a_1(t), a_2(t), \dots$, which have stochastic positions $v_1(t), v_2(t), \dots$. These methods are currently among the most successful and versatile for numerically solving the filtering problem. The basis of this class of numerical methods is the representation of π_t given by the Kallianpur–Striebel formula (see (2.2) below). In the case when the signal process is modelled by the solution of a stochastic differential equation (SDE) and the observation process is a function of the signal perturbed by white noise (see Section 2 below for further details), the formula entails the computation of expectations of functionals of the solution of the signal SDE that are parametrized by the observation path. The numerical approximation of π_t requires three procedures:

- the discretization of the functionals. The discretization corresponds to a choice of a partition of the time interval $[0, t]$.
- the approximation of the law of the signal with a discrete measure.
- the control of the computational effort.

The first step is typically achieved by the discretization scheme introduced by Picard in [10]. This offers a first order approximation for the functionals appearing in formula (2.2). More precisely, the L_1 -rate of convergence of the approximation is proportional with the mesh of the partition of the time interval $[0, t]$ (see Theorem 21.5 in [2]). The second and the third step are achieved by a combination of an Euler approximation of the solution of the SDE, a Monte Carlo step that gives a sample from the law of the Euler approximation and a re-sampling step that acts as a variance reduction method and keeps the computational effort in control. There are a variety of algorithms that follow this template. Further details can be found, for instance, in Part VII of [3]. It is worth pointing out that once the functional discretization and the Euler approximation have been applied, the problem can be reduced to one where the signal evolves and is observed in discrete time. The discrete version of the filtering problem is popular both with practitioners and with theoreticians. The majority of the existing theoretical results and the numerical algorithms are constructed and analyzed in the discrete framework. For more details, the interested reader can consult the comprehensive theoretical monograph [5] and the reference therein and the equally comprehensive methodological volume [6] and the references therein with some updates in Part VII of [3].

The first order discretization introduced by Picard creates a bottleneck: There exist higher order schemes for approximating the law of the signal that can

¹Also known as *particle filters* or *sequential Monte Carlo methods*.

be used, but which won't bring any substantial improvements because of this. For example, in the recent paper [4], the authors employ high order cubature methods to approximate the law of the signal with only minimal improvements due to the low order discretization of the required functionals. The aim of this paper is to address this issue. We introduce below second order discretization of the functionals. As we shall see, we prove that the L_p -rate of convergence of the approximation is proportional with the square of the mesh of the partition of the time interval $[0, t]$. For details, see Theorem 1 below. In a subsequent paper [9], this discretization procedure is employed to produce a second order particle filter. It is hoped that this discretization will be used in conjunction with other high order approximations of the law of the signal, in particular with cubature methods. It is worth mentioning we are not aware of any other similar discretization scheme and that, even though a class of schemes of any order would be desirable we haven't been able to construct one.²

The paper is organized as follows: In Section 3 we introduce some basic definitions and state the main result of the paper, Theorem 1. Section 3 is devoted to prove a general discretization result, Theorem 12, from which we will deduce our main result. In Section 5, we state some technical lemmas needed to apply Theorem 12 and we give the proof of Theorem 1. Finally, in Section 6 we give the proof of the technical lemmas introduced in the previous section.

2. The framework

Let (Ω, \mathcal{F}, P) be a probability space together with a filtration $(\mathcal{F}_t)_{t \geq 0}$ which satisfies the usual conditions. On (Ω, \mathcal{F}, P) we consider a $d_X \times d_Y$ -dimensional partially observed system (X, Y) satisfying

$$\begin{aligned} X_t &= X_0 + \int_0^t f(X_s) ds + \int_0^t \sigma(X_s) dV_s, \\ Y_t &= \int_0^t h(X_s) ds + W_t, \end{aligned}$$

where V is a standard \mathcal{F}_t -adapted d_V -dimensional Brownian motion and W is a standard \mathcal{F}_t -adapted d_Y -dimensional Brownian motion, independent of each other. We also denote by π_0 the law of X_0 . We assume that $f = (f^i)_{i=1}^{d_X} : \mathbb{R}^{d_X} \rightarrow \mathbb{R}^{d_X}$ and $\sigma = (\sigma_j^i)_{\substack{i=1, \dots, d_X, \\ j=1, \dots, d_V}} : \mathbb{R}^{d_X} \rightarrow \mathbb{R}^{d_X \times d_V}$ are globally Lipschitz

continuous. In addition, we assume that $h = (h^i)_{i=1}^{d_Y} : \mathbb{R}^{d_X} \rightarrow \mathbb{R}^{d_Y}$ is measurable and has linear growth.

Let $\{\mathcal{Y}_t\}_{t \geq 0}$ be the usual augmentation of the filtration associated with the process Y , that is, $\mathcal{Y}_t = \sigma(Y_s, s \in [0, t]) \vee \mathcal{N}$, where \mathcal{N} are all the P -null sets

²To be more precise, we can construct discretization schemes of any order, but not recursive ones. That means that the discretization at time t_1 , cannot be constructed by starting with the discretization at time $t_2 < t_1$ and adding to it the part corresponding to $[t_2, t_1]$. Instead we need to redo the discretization for the entire interval $[0, t]$, which will lead to a non-recursive particle filter. By contrast, the functional discretization presented here, as well as the original Picard discretisation, are recursive. See Remark 2 for further details.

of (Ω, \mathcal{F}, P) . We are interested in determining π_t , the conditional law of the signal X at time t given the information accumulated from observing Y in the interval $[0, t]$. More precisely, for any Borel measurable and bounded function φ , we want to compute $\pi_t(\varphi) = \mathbb{E}[\varphi(X_t) | \mathcal{Y}_t]$. By an application of Girsanov's theorem one can construct a new probability measure \tilde{P} absolutely continuous with respect to P under which Y becomes a Brownian motion independent of X in the law of X remains unchanged. Moreover the process $Z = (Z_t)_{t \geq 0}$ given by

$$Z_t = \exp \left(\sum_{i=1}^{d_Y} \int_0^t h^i(X_s) dY_s^i - \frac{1}{2} \sum_{i=1}^{d_Y} \int_0^t h^i(X_s)^2 ds \right), \quad t \geq 0. \quad (2.1)$$

is an \mathcal{F}_t -adapted martingale under \tilde{P} . Let $\tilde{\mathbb{E}}$ be the expectation with respect to \tilde{P} . In the following we will make use of the measure valued process $\rho = (\rho_t)_{t \geq 0}$, defined by the formula $\rho_t(\varphi) = \tilde{\mathbb{E}}[\varphi(X_t) Z_t | \mathcal{Y}_t]$, for any bounded Borel measurable function φ . The two processes π and ρ are connected through the Kallianpur-Striebel's formula:

$$\begin{aligned} \pi_t(\varphi) &= \frac{\rho_t(\varphi)}{\rho_t(\mathbf{1})} \\ &= \frac{\tilde{\mathbb{E}} \left[\varphi(X_t) \exp \left(\sum_{i=1}^{d_Y} \int_0^t h^i(X_s) dY_s^i - \frac{1}{2} \sum_{i=1}^{d_Y} \int_0^t h^i(X_s)^2 ds \right) \middle| \mathcal{Y}_t \right]}{\tilde{\mathbb{E}} \left[\exp \left(\sum_{i=1}^{d_Y} \int_0^t h^i(X_s) dY_s^i - \frac{1}{2} \sum_{i=1}^{d_Y} \int_0^t h^i(X_s)^2 ds \right) \middle| \mathcal{Y}_t \right]}, \end{aligned} \quad (2.2)$$

P-a.s., where $\mathbf{1}$ is the constant function $\mathbf{1}(x) = 1, x \in \mathbb{R}^d$. As a result, ρ is called the unnormalised conditional distribution of the signal. For further details on the filtering framework, see [1].

It follows from (2.2) that $\pi_t(\varphi)$ is a ratio of two conditional expectations of functionals of the signal that depend on the stochastic integrals with respect to the process Y . Hence, a second order discretization of π_t relies on the second order approximation of these two expectations. We achieve this in Theorem 1 below.

3. Main result

We introduce first some useful notation and definitions. We denote by \mathcal{B}_b the space of bounded Borel-measurable functions and by C_b^k the space of continuously differentiable functions up to order $k \in \mathbb{Z}_+$ with bounded derivatives of order greater or equal to one. Moreover, we denote by C_P^k the space of continuously differentiable functions up to order $k \in \mathbb{Z}_+$ such that the function and its derivatives have at most polynomial growth.

In the following, we will use the notation introduced in Section 5.4 in Kloeden and Platen [7]. More precisely, let S be a subset of \mathbb{Z}_+ and denote by $\mathcal{M}^*(S)$ the set of all multi-indices with values in S . In addition, denote by $\mathcal{M}(S) \triangleq$

$\mathcal{M}^*(S) \cup \{\emptyset\}$. For $\alpha = (\alpha_1, \dots, \alpha_k) \in \mathcal{M}(S)$ denote by $|\alpha| \triangleq k$ the length of α ($|\emptyset| = 0$), $\alpha_- \triangleq (\alpha_1, \dots, \alpha_{k-1})$ and $-\alpha \triangleq (\alpha_2, \dots, \alpha_k)$. Given two multi-indices $\alpha, \beta \in \mathcal{M}(S)$ we denote its concatenation by $\alpha * \beta$. We shall also consider the hierarchical set $\mathcal{M}_m(S)$ and its associated remainder set $\mathcal{M}_m^R(S)$, that is,

$$\mathcal{M}_m(S) \triangleq \{\alpha \in \mathcal{M}(S) : |\alpha| \leq m\}$$

and

$$\mathcal{M}_m^R(S) \triangleq \{\alpha \in \mathcal{M}(S) : |\alpha| = m + 1\}.$$

We shall use the sets of multi-indices with values in the sets $S_0 = \{0, 1, \dots, d_V\}$ and $S_1 = \{1, \dots, d_V\}$.

For $\alpha \in \mathcal{M}(S_0)$, denote by $I_\alpha(h)_{s,t}$ the following Itô iterated integral

$$I_\alpha(h(\cdot))_{s,t} = \begin{cases} h(t) & \text{if } |\alpha| = 0 \\ \int_s^t I_{\alpha_-}(h(\cdot))_{s,u} du & \text{if } |\alpha| \geq 1 \text{ and } \alpha_{|\alpha|} = 0 \\ \int_s^t I_{\alpha_-}(h(\cdot))_{s,u} dV_u^{\alpha_{|\alpha|}} & \text{if } |\alpha| \geq 1 \text{ and } \alpha_{|\alpha|} \neq 0 \end{cases},$$

where $h = \{h(t)\}_{t \in [0,T]}$ is an adapted process (satisfying appropriate integrability conditions) and $\alpha_0 = \emptyset$. For $\alpha \in \mathcal{M}(S_0)$, with $\alpha = (\alpha_1, \dots, \alpha_k)$, the differential operators L^α is defined by

$$\begin{aligned} L^\alpha g &= L^{\alpha_1} \circ L^{\alpha_2} \circ \dots \circ L^{\alpha_k} g, \\ L^\emptyset g &= g, \end{aligned}$$

where $L^0, L^r, r = 1, \dots, d_V$ are the differential operators defined by

$$\begin{aligned} L^0 g(x) &\triangleq \langle f(x), \nabla g(x) \rangle + \frac{1}{2} \sum_{k,l=1}^{d_X} (\sigma \sigma^T)_l^k(x) \frac{\partial^2 g}{\partial x^k \partial x^l}(x) \\ &= \sum_{k=1}^{d_X} f^k(x) \frac{\partial g}{\partial x^k}(x) + \frac{1}{2} \sum_{k,l=1}^{d_X} \sum_{r=1}^{d_V} \sigma_r^k(x) \sigma_r^l(x) \frac{\partial^2 g}{\partial x^k \partial x^l}(x). \\ L^r g(x) &\triangleq \langle \sigma_r(x), \nabla g(x) \rangle = \sum_{k=1}^{d_X} \sigma_r^k(x) \frac{\partial g}{\partial x^k}(x), \quad r = 1, \dots, d_V, \end{aligned}$$

where $g : \mathbb{R}^{d_X} \rightarrow \mathbb{R}$ belongs to $C_P^2(\mathbb{R}^{d_X}; \mathbb{R})$.

Let $\tau \triangleq \{0 = t_0 < \dots < t_i < \dots < t_n = t\}$ be a partition of $[0, t]$, $\delta_i \triangleq t_i - t_{i-1}$, $\delta \triangleq \max_{i=1, \dots, n} \delta_i$ and $\tau(s)$ is the largest element of the partition smaller than or equal to s , i.e., $\tau(s) \triangleq t_{i-1}, s \in [t_{i-1}, t_i], i = 1, \dots, n$. We denote by $\Pi(t)$ the set of all partitions of $[0, t]$ such that δ converges to zero when n tends to infinity and by $\Pi(t, \delta_0)$ the set of all partitions of $[0, t]$ such that δ converges to zero when n tends to infinity and $\delta < \delta_0$.

To simplify the notation, we will add an additional component to the Brownian motion Y . Let Y^0 be the process $Y_s^0 = s$, for all $s \geq 0$ and consider the

$(d_Y + 1)$ -dimensional process $Y = (Y^i)_{i=0}^{d_Y}$. Then the martingale Z defined in (2.1) can be written as

$$Z_t = \exp \left(\sum_{i=1}^{d_Y} \int_0^t h^i(X_s) dY_s^i \right), \quad t \geq 0,$$

where $h^0 = -\frac{1}{2} \sum_{i=1}^{d_Y} (h^i)^2$. For $\tau \in \Pi(t)$, consider the process $Z^{\tau,2} = (Z_t^{\tau,2})_{t \geq 0}$ given by

$$\begin{aligned} Z_t^{\tau,2} &= \prod_{j=0}^{n-1} \exp \left(\sum_{i=0}^{d_Y} h^i(X_{t_j}) (Y_{t_{j+1}} - Y_{t_j}) + L^0 h^i(X_{t_j}) \int_{t_j}^{t_{j+1}} (s - t_j) dY_s^i \right. \\ &\quad \left. + L^r h^i(X_{t_j}) \int_{t_j}^{t_{j+1}} (V_s^r - V_{\tau(s)}^r) dY_s^i \right) \\ &= \exp \left(\sum_{i=0}^{d_Y} \int_0^t \left(h^i(X_{\tau(s)}) + L^0 h^i(X_{\tau(s)})(s - \tau(s)) + \sum_{r=1}^{d_Y} L^r h^i(X_{\tau(s)})(V_s^r - V_{\tau(s)}^r) \right) dY_s^i \right) \end{aligned} \quad (3.1)$$

In the following, we will use the standard notation $L^p(\Omega, \mathcal{F}, \tilde{P})$ for the space of p -integrable random variables (with respect to \tilde{P}) and denote by $\|\cdot\|_p$, the corresponding norm on $L^p(\Omega, \mathcal{F}, \tilde{P})$, i.e., for $\xi \in L^p(\Omega, \mathcal{F}, \tilde{P})$, $\|\xi\|_p \triangleq \tilde{\mathbb{E}}[|\xi|^p]^{1/p}$. For any Borel measurable function φ such that $\varphi(X_t) Z_t^{\tau,2} \in L^1(\Omega, \mathcal{F}, \tilde{P})$ define

$$\begin{aligned} \rho_t^{\tau,2}(\varphi) &\triangleq \tilde{\mathbb{E}}[\varphi(X_t) Z_t^{\tau,2} | \mathcal{Y}_t], \\ \pi_t^{\tau,2}(\varphi) &\triangleq \rho_t^{\tau,2}(\varphi) / \rho_t^{\tau,2}(\mathbf{1}). \end{aligned}$$

Our main result is the following:

Theorem 1. *Suppose that $f, \sigma \in \mathcal{B}_b \cap C_b^2$, $h \in \mathcal{B}_b \cap C_b^2 \cap C_P^4$ and that X_0 has moments of all orders. Then, for any $p \geq 1$ and $\varphi \in C_P^2$ there exists a constant $C = C(t, p, \varphi)$ independent of $\tau \in \Pi(t, \delta_0)$, where*

$$\delta_0 = \frac{1}{2p \|Lh\|_\infty \sqrt{d_Y d_V}}, \quad (3.2)$$

such that

$$\left\| \rho_t(\varphi) - \rho_t^{\tau,2}(\varphi) \right\|_p \leq C \delta^2.$$

Moreover, if $\sup_{\tau \in \Pi(t, \delta_0)} \left\| \pi_t^{\tau,2}(\varphi) \right\|_{2p+\varepsilon} < \infty$, for some $\varepsilon > 0$, then

$$\left\| \pi_t(\varphi) - \pi_t^{\tau,2}(\varphi) \right\|_p \mathbb{E}[\left| \pi_t(\varphi) - \pi_t^{\tau,2}(\varphi) \right|^p] \leq C \delta^{2p},$$

where C is another constant independent of $\tau \in \Pi(t, \delta_0)$.

Remark 2. *i. The functional discretization given in (3.1) is recursive. More precisely, if $\tau' \in \Pi(t+s)$ is a partition that includes t as an intermediate point, for example $\tau' \triangleq \{0 = t_0 < \dots < t_k = t < t_{k+1} < \dots < t_n = t+s\}$ with $0 < k < n$, then*

$$Z_{t+s}^{\tau',2} = Z_t^{\tau,2} \prod_{k=i}^{n-1} \exp \left(\sum_{i=0}^{d_Y} h^i(X_{t_k}) (Y_{t_{k+1}} - Y_{t_k}) + L^0 h^i(X_{t_k}) \int_{t_k}^{t_{k+1}} (s - t_k) dY_s^i + L^r h^i(X_{t_k}) \int_{t_k}^{t_{k+1}} (V_s^r - V_{\tau(s)}^r) dY_s^i \right).$$

This property is essential for implementation purposes as at every discretization time we only need to use the previous functional discretization and the term corresponding to the next interval to obtain the new functional discretization.

ii. The second order discretization presented above is obtained by making use of the first order Itô-Taylor expansion of $h^i(X_s)$ $i = 0, \dots, d_Y$. Of course one can generalize this in the following straightforward manner. Let $\xi^{\tau,m} = (\xi_i^{\tau,m})_{i=0}^{d_Y}$, $m \in \mathbb{N}$ be the random vectors obtained by using an $(m-1)$ -order Itô-Taylor expansion of $h^i(X_s)$, more precisely

$$\begin{aligned} \xi_i^{\tau,m} &\triangleq \sum_{j=1}^n \sum_{\alpha \in \mathcal{M}_{m-1}(S_0)} L^\alpha h^i(X_{t_{j-1}}) \int_{t_{j-1}}^{t_j} I_\alpha(\mathbf{1})_{t_{j-1},s} dY_s^i \\ &= \sum_{\alpha \in \mathcal{M}_{m-1}(S_0)} \int_0^t L^\alpha h^i(X_{\tau(s)}) I_\alpha(\mathbf{1})_{\tau(s),s} dY_s^i. \end{aligned}$$

Using this notation we can write

$$\rho_t^{\tau,2}(\varphi) = \tilde{\mathbb{E}} \left[\varphi(X_t) \exp \left(\sum_{i=0}^{d_Y} \xi_i^{\tau,2} \right) \middle| \mathcal{Y}_t \right].$$

As an immediate generalization, we could replace $\xi_i^{\tau,2}$ with $\xi_i^{\tau,m}$ to obtain an m -order discretization of $\rho_t^\tau(\varphi)$. Unfortunately this is not possible as $\xi_i^{\tau,m}$ does not have finite exponential moments for $m \geq 3$.

iii. A non-recursive m -order functional discretization can be constructed, as follows

$$\rho_t^{\tau,2}(\varphi) = \tilde{\mathbb{E}} \left[\varphi(X_t) \exp \left(\Psi^m \left(\sum_{i=0}^{d_Y} \xi_i^{\tau,m} \right) \right) \middle| \mathcal{Y}_t \right],$$

where Ψ^m is a suitably chosen truncation function. This result is an immediate Corollary of Theorem 12 below.

4. A general approximation result

We will not prove Theorem 1 directly. Instead, we will first show a more general approximation result and we will deduce Theorem 1 as a consequence. We start

by introducing some technical conditions and recalling some basic results on martingale representations.

Condition 3 ($\mathbf{S}(m)$). *All moments of X_0 are finite. The functions $f = (f^i)_{i=1}^{d_X} : \mathbb{R}^{d_X} \rightarrow \mathbb{R}^{d_X}$, $\sigma = (\sigma_j^i)_{i=1, \dots, d_X, j=1, \dots, d_V} : \mathbb{R}^{d_X} \rightarrow \mathbb{R}^{d_X \times d_V}$ belong to C_P^m and are globally Lipschitz.*

Note that if condition $\mathbf{S}(m)$ holds for some $m \in \mathbb{N}$, then condition $\mathbf{S}(n)$ holds for any $n \leq m$.

Remark 4. *Under condition $\mathbf{S}(m)$, in particular if the coefficients are globally Lipschitz and all moments of X_0 are finite, the signal process X has moments of all orders and for any $p > 0$, we have*

$$\tilde{\mathbb{E}} \left[\sup_{s \in [0, t]} |X_s|^p \right] < \infty.$$

Following the notation from the previous section, let $\xi = (\xi_i)_{i=0}^{d_Y}$ be the random vector with entries

$$\xi_i = \int_0^t g^i(X_s) dY_s^i, \quad i = 0, \dots, d_Y,$$

where $g : \mathbb{R}^{d_X} \rightarrow \mathbb{R}^{d_Y+1}$. For the remainder of the section we assume that g satisfies the following regularity assumption:

Condition 5 ($\mathbf{G}(m)$). *The function $g : \mathbb{R}^{d_X} \rightarrow \mathbb{R}^{d_Y+1}$ is a C_P^{2m} function.*

Next, let $\xi^{\tau, m} = (\xi_i^{\tau, m})_{i=0}^{d_Y}$, $m \in \mathbb{N}$ be the random vectors with entries

$$\begin{aligned} \xi_i^{\tau, m} &\triangleq \sum_{j=1}^n \sum_{\alpha \in \mathcal{M}_{m-1}(S_0)} L^\alpha g^i(X_{t_{j-1}}) \int_{t_{j-1}}^{t_j} I_\alpha(\mathbf{1})_{t_{j-1}, s} dY_s^i \\ &= \sum_{\alpha \in \mathcal{M}_{m-1}(S_0)} \int_0^t L^\alpha g^i(X_{\tau(s)}) I_\alpha(\mathbf{1})_{\tau(s), s} dY_s^i. \end{aligned}$$

Let $\varphi : \mathbb{R}^{d_X} \rightarrow \mathbb{R}$ be a measurable function and $\psi : \mathbb{R}^{d_Y+1} \rightarrow \mathbb{R}$ be a continuously differentiable function. We are interested in finding high order bounds of the following quantity, called henceforth the approximation error,

$$q(X_t, \xi, \xi^{\tau, m}) \triangleq \tilde{\mathbb{E}}[\varphi(X_t)(\psi(\xi) - \psi(\xi^{\tau, m})) | \mathcal{Y}_t],$$

in terms of δ , the size of the partition τ . Note that, by the mean value theorem we can write

$$\varphi(X_t)(\psi(\xi) - \psi(\xi^{\tau, m})) = \sum_{i=0}^{d_Y} \eta_i^{\tau, m} (\xi_i - \xi_i^{\tau, m}), \quad (4.1)$$

where $\eta^{\tau,m} = (\eta_i^{\tau,m})_{i=0}^{d_Y}$ is the random vector with components

$$\eta_i^{\tau,m} = \int_0^1 \varphi(X_t) \partial_i \psi(s\xi + (1-s)\xi^{\tau,m}) ds. \quad (4.2)$$

Then, it is natural to consider the following set of conditions parametrized by $p \geq 1, m \in \mathbb{N}$ and Π a set of partitions:

Condition 6 ($\mathbf{L}(p, m, \Pi)$). *There exists $\varepsilon > 0$ such that*

$$\sup_{\tau \in \Pi} \sup_{s \in [0,1]} \tilde{\mathbb{E}}[|\varphi(X_t) \partial_i \psi(s\xi + (1-s)\xi^{\tau,m})|^{2p+\varepsilon}] < \infty, \quad (4.3)$$

for all $i = 0, \dots, d_Y$.

Remark 7. *Note that ξ has moments of all orders. As the functions $L^\alpha g^i(x), i = 0, \dots, d_Y, \alpha \in \mathcal{M}_m(S_0)$ have polynomial growth, then $\xi^{\tau,m}$ also has moments of all orders. If the function φ and the partial derivatives of ψ have, at most, polynomial growth, then condition $\mathbf{L}(p, m, \Pi)$ is satisfied.*

Note, however, that in the filtering problem the function ψ is the exponential and the previous remark will not apply. A different approach will be required in the case to show that condition $\mathbf{L}(p, m, \Pi)$ is satisfied.

Theorem 8. *Assume condition $\mathbf{L}(p, m, \Pi)$ is satisfied. Then there exists $\varepsilon > 0$ such that the random variables $\eta_i^{\tau,m} \in L^{2p+\varepsilon}(\Omega, \mathcal{F}, \tilde{P}), i = 0, \dots, d_Y$, and admit the following martingale representation*

$$\eta_i^{\tau,m} = \hat{\eta}_i^{\tau,m} + \sum_{r_1=1}^{d_V} \int_0^t \psi_{i,r_1}^{\tau,m}(s_1) dV_{s_1}^{r_1}, \quad i = 0, \dots, d_Y,$$

where $\hat{\eta}_i^{\tau,m} = \tilde{\mathbb{E}}[\eta_i^{\tau,m} | \mathcal{F}_0 \vee \mathcal{Y}_t]$, belong to $L^{2p+\varepsilon}(\Omega, \mathcal{F}_0 \vee \mathcal{Y}_t, \tilde{P}), i = 0, \dots, d_Y$. Moreover, $\psi_{i,r_1}^{\tau,m} = \{\psi_{i,r_1}^{\tau,m}(s_1), s_1 \in [0, t]\}$ are progressively measurable $\mathcal{F}_{s_1} \vee \mathcal{Y}_t$ -adapted processes such that

$$\tilde{\mathbb{E}} \left[\left(\int_0^t \psi_{i,r_1}^{\tau,m}(s_1)^2 ds_1 \right)^{(2p+\varepsilon)/2} \right] < \infty,$$

for all $i = 0, \dots, d_Y$ and $r_1 = 1, \dots, d_V$.

By iterating the integral representation in Theorem 8, one can get the following result.

Theorem 9 (Stroock-Taylor). *Assume condition $\mathbf{L}(p, m, \Pi)$, then for any $k \in \mathbb{N}$ the random variables $\eta_i^{\tau,m}$ admit the following integral representation*

$$\eta_i^{\tau,m} = \sum_{\beta \in \mathcal{M}_k(S_1)} \hat{I}_\beta(\hat{\psi}_{i,\beta}^{\tau,m}(\cdot))_{0,t} + \sum_{\beta \in \mathcal{M}_k^R(S_1)} \hat{I}_\beta(\psi_{i,\beta}^{\tau,m}(\cdot))_{0,t},$$

where the kernels $\hat{\psi}_{i,\beta}^{\tau,m}$ and $\psi_{i,\beta}^{\tau,m}$ satisfy the following recursive relationship

$$\hat{\psi}_{i,\emptyset}^{\tau,m} \triangleq \tilde{\mathbb{E}}[\eta_i^{\tau,m} | \mathcal{F}_0 \vee \mathcal{Y}_t],$$

$$\hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|}) \triangleq \tilde{\mathbb{E}} \left[\psi_{i,\beta}^{\tau,m}(\cdot) | \mathcal{F}_0 \vee \mathcal{Y}_t \right], \quad 1 \leq |\beta| \leq k$$

and

$$\begin{aligned} \eta_i^{\tau,m} &= \hat{\psi}_{i,\emptyset}^{\tau,m} + \sum_{r_1=1}^{d_V} \int_0^t \psi_{i,\beta_1}^{\tau,m}(s_1) dV_{s_1}^{\beta_1}, \\ \psi_{i,\beta_-}^{\tau,m}(s_1, \dots, s_{|\beta|-1}) &= \hat{\psi}_{i,\beta_-}^{\tau,m}(s_1, \dots, s_{|\beta|-1}) \\ &\quad + \sum_{\beta_{|\beta|=1}}^{d_V} \int_0^{s_{|\beta|-1}} \psi_{i,\beta_- * (\beta_{|\beta|})}^{\tau,m}(s_1, \dots, s_{|\beta|-1}, s_{|\beta|}) dV_{s_{|\beta|}}^{\beta_{|\beta|}}. \end{aligned}$$

If $\eta_i^{\tau,m}$ is Malliavin differentiable up to order $k+1$ then

$$\hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|}) = \tilde{\mathbb{E}} \left[D_{s_1, \dots, s_{|\beta|}}^{\beta_1, \dots, \beta_{|\beta|}} \eta_i^{\tau,m} | \mathcal{F}_0 \vee \mathcal{Y}_t \right], \quad 0 \leq |\beta| \leq k,$$

and

$$\psi_{i,\beta}^{\tau,m}(s_1, \dots, s_{k+1}) = \tilde{\mathbb{E}} \left[D_{s_1, \dots, s_{k+1}}^{\beta_1, \dots, \beta_{k+1}} \eta_i^{\tau,m} | \mathcal{F}_{s_{k+1}} \vee \mathcal{Y}_t \right], \quad |\beta| = k+1.$$

Remark 10. For any $\beta \in \mathcal{M}_k(S_1)$, $|\beta| \geq 1$, we have that $\hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|})$ is $\mathcal{F}_0 \vee \mathcal{Y}_t$ measurable kernel defined on the simplex

$$\mathcal{S}_{|\beta|}(t) \triangleq \{(s_1, \dots, s_{|\beta|}) \in [0, t]^{|\beta|} : 0 \leq s_{|\beta|} < \dots < s_1 \leq t\},$$

and

$$\hat{I}_\beta(\hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|}))_{0,t} \triangleq \int_0^t \int_0^{s_1} \dots \int_0^{s_{|\beta|-1}} \hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|}) dV_{s_{|\beta|}}^{\beta_{|\beta|}} \dots dV_{s_1}^{\beta_1}.$$

If $\beta \in \mathcal{M}_k^R(S_1)$ the kernel $\psi_{i,\beta}^{\tau,m}(s_1, \dots, s_{k+1})$ is a $\mathcal{F}_{s_{k+1}} \vee \mathcal{Y}_t$ adapted process and

$$\hat{I}_\beta(\psi_{i,\beta}^{\tau,m}(s_1, \dots, s_{k+1}))_{0,t} = \int_0^t \int_0^{s_1} \dots \int_0^{s_k} \psi_{i,\beta}^{\tau,m}(s_1, \dots, s_{k+1}) dV_{s_{k+1}}^{\beta_{k+1}} \dots dV_{s_1}^{\beta_1}.$$

We also consider the following set of conditions parametrized by $p \geq 1, m \in \mathbb{N}$ and Π a set of partitions.

Condition 11 (UK(p, m, Π)). There exists $\varepsilon > 0$ such that the kernels $\{\hat{\psi}_{i,\beta}^{\tau,m}(\cdot)\}_{\beta \in \mathcal{M}_{m-1}(S_1)}$ and $\{\psi_{i,\beta}^{\tau,m}(\cdot)\}_{\beta \in \mathcal{M}_{m-1}^R(S_1)}$, given in Theorem 9, satisfy

$$\sup_{\tau \in \Pi} \sup_{0 \leq s_{|\beta|} < \dots < s_1 \leq t} \tilde{\mathbb{E}} \left[\left| \hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|}) \right|^{2p \vee (2+\varepsilon)} \right] < \infty,$$

and

$$\sup_{\tau \in \Pi} \sup_{0 \leq s_m < \dots < s_1 \leq t} \tilde{\mathbb{E}} \left[\left| \psi_{i,\beta}^{\tau,m}(s_1, \dots, s_m) \right|^{2p \vee (2+\varepsilon)} \right] < \infty,$$

for all $i = 0, \dots, d_Y$.

The following theorem gives the general discretisation error that will allow us to deduce Theorem 1.

Theorem 12. *Assume that conditions $\mathbf{S}(m)$, $\mathbf{G}(m)$, $\mathbf{L}(p, m, \Pi)$ and $\mathbf{UK}(p, m, \Pi)$ hold. Then, there exists a constant $C = C(t)$ independent of the partition $\tau \in \Pi$ such that*

$$\|q(X_t, \xi, \xi^{\tau, m})\|_{2p} \leq C\delta^m.$$

Proof. We can write

$$\begin{aligned} \xi_i - \xi_i^{\tau, m} &= \int_0^t \{g^i(X_s) - \sum_{\alpha \in \mathcal{A}_{m-1}} \int_0^t L^\alpha g^i(X_{\tau(s)}) I_\alpha(\mathbf{1})_{\tau(s), s}\} dY_s^i \\ &\triangleq \int_0^t \Theta_s^{g^i, \tau, m} dY_s^i. \end{aligned} \quad (4.4)$$

Next, by the Itô-Taylor expansion with hierarchical set $\mathcal{M}_{m-1}(S_0)$, see Theorem 5.5.1 in Kloeden-Platen [7], we have that

$$\Theta_s^{g^i, \tau, m} = \sum_{\alpha \in \mathcal{M}_{m-1}^R(S_0)} I_\alpha(L^\alpha g^i(X_{\cdot}))_{\tau(s), s}.$$

Assumptions $\mathbf{S}(m)$ and $\mathbf{G}(m)$ imply the polynomial growth of $L^\alpha g^i, \alpha \in \mathcal{M}_{m-1}^R(S_0), i = 1, \dots, d_Y$. In addition, one gets that

$$\begin{aligned} &\tilde{\mathbb{E}}[\xi_i - \xi_i^{\tau, m} | \mathcal{F}_0 \vee \mathcal{Y}_t] \\ &= \tilde{\mathbb{E}} \left[\int_0^t \int_{\tau(s)}^s \int_{\tau(s)}^{s_{m-1}} \cdots \int_{\tau(s)}^{s_1} L^{\alpha_0^m} g^i(X_{s_0}) ds_0 \cdots ds_{m-1} dY_s^i | \mathcal{F}_0 \vee \mathcal{Y}_t \right] \\ &= \int_0^t \int_{\tau(s)}^s \int_{\tau(s)}^{s_{m-1}} \cdots \int_{\tau(s)}^{s_1} \tilde{\mathbb{E}}[L^{\alpha_0^m} g^i(X_{s_0}) | \mathcal{F}_0 \vee \mathcal{Y}_t] ds_0 \cdots ds_{m-1} dY_s^i, \end{aligned}$$

where $\alpha_0^m = \overbrace{(0, \dots, 0)}^m$ and if $m = 1$ then the integral is just over the simplex $\{(s_0, s) : \tau(s) \leq s_0 \leq s, 0 \leq s \leq t\}$. Note that

$$\tilde{\mathbb{E}}[L^{\alpha_0^m} g^i(X_{s_0}) | \mathcal{F}_0 \vee \mathcal{Y}_t] = \tilde{\mathbb{E}}[L^{\alpha_0^m} g^i(X_{s_0}) | \mathcal{F}_0] = P_{s_0} L^{\alpha_0^m} g^i(X_0),$$

where $P_t g(x) \triangleq \tilde{\mathbb{E}}[g(X_t^x)]$ is the semigroup associated to the signal, that is, to the SDE

$$X_t^x = x + \int_0^t f(X_s^x) ds + \int_0^t \sigma(X_s^x) dV_s.$$

Moreover, under the assumption $\mathbf{S}(m)$ the following bound holds. For any $p \geq 2$,

$$\tilde{\mathbb{E}} \left[\sup_{s \in [0, t]} |X_s^x|^p \right] \leq C(p, t)(1 + |x|^p).$$

Hence, as $|L^{\alpha_0^m} g^i(x)| \leq C(g^i, f, \sigma)(1 + |x|^r)$ for some $r \geq 2$, we have that

$$\begin{aligned} P_{s_0} L^{\alpha_0^m} g^i(X_0) &= \mathbb{E}[L^{\alpha_0^m} g^i(X_{s_0}^x)]|_{x=X_0} \\ &\leq \mathbb{E}[C(g^i, f, \sigma)(1 + |X_{s_0}^x|^r)]|_{x=X_0} \\ &\leq C(g^i, f, \sigma) \left(1 + \mathbb{E}[\sup_{0 \leq s_0 \leq t} |X_{s_0}^x|^r]|_{x=X_0}\right) \\ &\leq C(g^i, f, \sigma, r, t)(1 + |X_0|^r). \end{aligned}$$

Taking into account equation (4.4) and using Theorem 9 with $k = m - 1$, we can write

$$\begin{aligned} q(X_t, \xi, \xi^{\tau, m}) &= \sum_{i=0}^{d_Y} \tilde{\mathbb{E}}[\hat{\eta}_i^{\tau, m}(\xi_i - \xi_i^{\tau, m}) | \mathcal{Y}_t] \\ &\quad + \sum_{i=0}^{d_Y} \sum_{\beta \in \mathcal{M}_{m-1}(S_1)} \tilde{\mathbb{E}}[\hat{I}_\beta(\hat{\psi}_{i, \beta}^{\tau, m}(\cdot))_{0, t}(\xi_i - \xi_i^{\tau, m}) | \mathcal{Y}_t] \\ &\quad + \sum_{i=0}^{d_Y} \sum_{\beta \in \mathcal{M}_{m-1}^R(S_1)} \tilde{\mathbb{E}}[\hat{I}_\beta(\psi_{i, \beta}^{\tau, m}(\cdot))_{0, t}(\xi_i - \xi_i^{\tau, m}) | \mathcal{Y}_t] \\ &\triangleq \sum_{i=0}^{d_Y} A_{i,1} + A_{i,2} + A_{i,3}. \end{aligned}$$

To finish the proof we will show that $\tilde{\mathbb{E}}[A_{i,j}^2] \leq C\delta^{2m}$, $i = 0, \dots, d_Y$, $j = 1, \dots, 3$, for some constant C that does not depend on the partition τ .

Term $A_{i,1}$:

For any $i = 1, \dots, d_Y$ and $\varepsilon > 0$ as in condition $\mathbf{L}(p, m, \Pi)$, we have that

$$\begin{aligned} &\tilde{\mathbb{E}}[A_{i,1}^{2p}] \\ &= \tilde{\mathbb{E}} \left[\tilde{\mathbb{E}} \left[\hat{\eta}_i^{\tau, m} \tilde{\mathbb{E}}[(\xi_i - \xi_i^{\tau, m}) | \mathcal{F}_0 \vee \mathcal{Y}_t] | \mathcal{Y}_t \right]^{2p} \right] \\ &\leq \tilde{\mathbb{E}} \left[(\hat{\eta}_i^{\tau, m})^{2p} \left(\int_0^t \int_{\tau(s)}^s \int_{\tau(s)}^{s_{m-1}} \cdots \int_{\tau(s)}^{s_1} \tilde{\mathbb{E}}[L^{\alpha_0^m} g^i(X_{s_0}) | \mathcal{F}_0 \vee \mathcal{Y}_t] ds_0 \cdots ds_{m-1} dY_s^i \right)^{2p} \right], \\ &\leq \tilde{\mathbb{E}} \left[(\hat{\eta}_i^{\tau, m})^{2p+\varepsilon} \right]^{\frac{2p}{2p+\varepsilon}} \\ &\quad \times \tilde{\mathbb{E}} \left[\left(\int_0^t \int_{\tau(s)}^s \int_{\tau(s)}^{s_{m-1}} \cdots \int_{\tau(s)}^{s_1} \tilde{\mathbb{E}}[L^{\alpha_0^m} g^i(X_{s_0}) | \mathcal{F}_0 \vee \mathcal{Y}_t] ds_0 \cdots ds_{m-1} dY_s^i \right)^{2p \frac{2p+\varepsilon}{\varepsilon}} \right]^{\frac{\varepsilon}{2p+\varepsilon}} \\ &\leq \|\hat{\eta}_i^{\tau, m}\|_{L^{2p+\varepsilon}(\tilde{P})}^{2p} \\ &\quad \times \tilde{\mathbb{E}} \left[\left(\int_0^t \left(\int_{\tau(s)}^s \int_{\tau(s)}^{s_{m-1}} \cdots \int_{\tau(s)}^{s_1} C(1 + |X_0|^r) ds_0 \cdots ds_{m-1} \right)^2 ds \right)^{p \frac{2p+\varepsilon}{\varepsilon}} \right]^{\frac{\varepsilon}{2p+\varepsilon}} \end{aligned}$$

$$\begin{aligned}
&\leq \|\hat{\eta}_i^{\tau,m}\|_{L^{2p+\varepsilon}(\tilde{P})}^{2p} \tilde{\mathbb{E}} \left[\left(\int_0^t C^2 (1 + |X_0|^r)^2 \delta^{2m} ds \right)^{p \frac{2p+\varepsilon}{\varepsilon}} \right]^{\frac{\varepsilon}{2p+\varepsilon}} \\
&\leq C^{2p} t^p \|\hat{\eta}_i^{\tau,m}\|_{L^p(\tilde{P})}^2 \tilde{\mathbb{E}} \left[(1 + |X_0|^r)^{2p \frac{2p+\varepsilon}{\varepsilon}} \right]^{\frac{\varepsilon}{2p+\varepsilon}} \delta^{2pm}.
\end{aligned}$$

Term $A_{i,2}$:

For any $i = 1, \dots, d_Y$, $\beta = (\beta_1, \dots, \beta_{|\beta|}) \in \mathcal{M}_{m-1}(S_1)$ and $\varepsilon > 0$ as in conditions **L**(p, m, Π) and **UK**(p, m, Π), we can use integration by parts to obtain

$$\begin{aligned}
(\xi_i - \xi_i^{\tau,m}) \hat{I}_\beta(\hat{\psi}_{i,\beta}^{\tau,m}(\cdot))_{0,t} &= \int_0^t \Theta_s^{g^i, \tau, m} dY_s^i \int_0^t \hat{I}_{-\beta}(\hat{\psi}_{i,\beta}^{\tau,m}(s, \cdot))_{0,s} dV_s^{\beta_1} \\
&= \int_0^t \left(\int_0^s \Theta_u^{g^i, \tau, m} dY_u^i \right) \hat{I}_{-\beta}(\hat{\psi}_{i,\beta}^{\tau,m}(s, \cdot))_{0,s} dV_s^{\beta_1} \\
&\quad + \int_0^t \hat{I}_\beta(\hat{\psi}_{i,\beta}^{\tau,m}(\cdot))_{0,s} \Theta_s^{g^i, \tau, m} dY_s^i.
\end{aligned}$$

Note that

$$\begin{aligned}
&\tilde{\mathbb{E}} \left[\left(\int_0^t \left(\int_0^s \Theta_u^{g^i, \tau, m} dY_u^i \right) \hat{I}_{-\beta}(\hat{\psi}_{i,\beta}^{\tau,m}(s, \cdot))_{0,s} dV_s^{\beta_1} \right)^2 \right] \\
&= \int_0^t \tilde{\mathbb{E}} \left[\left(\left(\int_0^s \Theta_u^{g^i, \tau, m} dY_u^i \right) \hat{I}_{-\beta}(\hat{\psi}_{i,\beta}^{\tau,m}(s, \cdot))_{0,s} \right)^2 \right] ds \\
&\leq \int_0^t \tilde{\mathbb{E}} \left[\left(\int_0^s \Theta_u^{g^i, \tau, m} dY_u^i \right)^{\frac{2(2+\varepsilon)}{\varepsilon}} \right]^{\varepsilon/(2+\varepsilon)} \tilde{\mathbb{E}} \left[\left(\hat{I}_{-\beta}(\hat{\psi}_{i,\beta}^{\tau,m}(s, \cdot))_{0,s} \right)^{2+\varepsilon} \right]^{2/(2+\varepsilon)} ds \\
&\leq \tilde{\mathbb{E}} \left[\sup_{0 \leq s \leq t} \left(\int_0^s \Theta_u^{g^i, \tau, m} dY_u^i \right)^{\frac{2(2+\varepsilon)}{\varepsilon}} \right]^{\varepsilon/(2+\varepsilon)} \int_0^t \tilde{\mathbb{E}} \left[\left(\hat{I}_{-\beta}(\hat{\psi}_{i,\beta}^{\tau,m}(s, \cdot))_{0,s} \right)^{2+\varepsilon} \right]^{2/(2+\varepsilon)} ds \\
&\leq C(\varepsilon, t) \tilde{\mathbb{E}} \left[\left(\int_0^t |\Theta_u^{g^i, \tau, m}|^2 du \right)^{\frac{2+\varepsilon}{\varepsilon}} \right]^{\varepsilon/(2+\varepsilon)} \int_0^t \tilde{\mathbb{E}} \left[\left(\hat{I}_{-\beta}(\hat{\psi}_{i,\beta}^{\tau,m}(s, \cdot))_{0,s} \right)^{2+\varepsilon} \right]^{2/(2+\varepsilon)} ds \\
&\leq C(\varepsilon, t) \tilde{\mathbb{E}} \left[\sup_{0 \leq u \leq t} |\Theta_u^{g^i, \tau, m}|^{\frac{2(2+\varepsilon)}{\varepsilon}} \right]^{\varepsilon/(2+\varepsilon)} \\
&\quad \times \left(\sup_{0 \leq s_{|\beta|} < \dots < s_1 \leq t} \tilde{\mathbb{E}} \left[\left| \hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|}) \right|^{2+\varepsilon} \right] \right)^{2/(2+\varepsilon)} \\
&< \infty,
\end{aligned}$$

where we have used the Burkholder-Davis-Gundy inequality several times, assumption **UK**(p, m, Π) and that $\sup_{0 \leq u \leq t} |\Theta_u^{g^i, \tau, m}|$ has moments of all orders.

This yields that

$$\int_0^v \left(\int_0^s \Theta_u^{g^i, \tau, m} dY_u^i \right) \hat{I}_{-\beta}(\hat{\psi}_{i, \beta}^{\tau, m}(s, \cdot))_{0, s} dV_s^{\beta_1}, \quad v \in [0, t],$$

is a $\mathcal{F}_v \vee \mathcal{Y}_t$ -martingale and it vanishes when taking conditional expectation with respect to $\mathcal{F}_0 \vee \mathcal{Y}_t$. Therefore,

$$\begin{aligned} & \tilde{\mathbb{E}} \left[(\xi_i - \xi_i^{\tau, m}) \hat{I}_{\beta}(\hat{\psi}_{i, \beta}^{\tau, m}(\cdot))_{0, t} | \mathcal{Y}_t \right] \\ &= \int_0^t \tilde{\mathbb{E}} \left[\hat{I}_{\beta}(\hat{\psi}_{i, \beta}^{\tau, m}(\cdot))_{0, s} \Theta_s^{g^i, \tau, m} | \mathcal{Y}_t \right] dY_s^i \\ &= \int_0^t \tilde{\mathbb{E}} \left[\int_0^s \cdots \int_0^{s_{|\beta|-1}} \hat{\psi}_{i, \beta}^{\tau, m}(s_1, \dots, s_{|\beta|}) dV_{s_{|\beta|}}^{\beta_{|\beta|}} \cdots dV_{s_1}^{\beta_1} \right. \\ &\quad \times \left. \left(\sum_{\alpha \in \mathcal{M}_{m-1}^R(S_0)} I_{\alpha}(L^{\alpha} g^i(X))_{\tau(s), s} \right) | \mathcal{Y}_t \right] dY_s^i \\ &= \int_0^t \tilde{\mathbb{E}} \left[\int_0^s \cdots \int_0^{s_{|\beta|-1}} \hat{\psi}_{i, \beta}^{\tau, m}(s_1, \dots, s_{|\beta|}) dV_{s_{|\beta|}}^{\beta_{|\beta|}} \cdots dV_{s_1}^{\beta_1} \right. \\ &\quad \times \left. \left(\sum_{\alpha \in \mathcal{M}_{m-1}^R(S_0)} \int_0^s \cdots \int_0^{s_2} \prod_{j=1}^m \mathbf{1}_{\{s_j > \tau(s)\}} L^{\alpha} g^i(X_{s_1}) dV_{s_1}^{\alpha_1} \cdots dV_{s_m}^{\alpha_m} \right) | \mathcal{Y}_t \right] dY_s^i, \end{aligned}$$

where $dV_{s_j}^{\alpha_j} = ds_j$ if $\alpha_j = 0$. Taking conditional expectation with respect to $\mathcal{F}_0 \vee \mathcal{Y}_t$ we get that the only term that does not vanish is the one corresponding to $\alpha = \alpha(\beta) \triangleq \alpha_0^{m-|\beta|} * (\beta_{|\beta|}, \dots, \beta_1)$. Hence, defining

$$\Lambda_{\alpha(\beta)}(s_{|\beta|}) \triangleq \int_{\tau(s)}^{s_{|\beta|}} \int_{\tau(s)}^{s_{|\beta|+1}} \cdots \int_{\tau(s)}^{s_{m-1}} L^{\alpha(\beta)} g^i(X_{s_m}) ds_m \cdots ds_{|\beta|+1},$$

we get that

$$\begin{aligned} & \tilde{\mathbb{E}} [\Lambda_{\alpha(\beta)}(s_{|\beta|})^2 | \mathcal{Y}_t] \\ & \leq \delta^{m-|\beta|} \int_{\tau(s)}^{s_{|\beta|}} \int_{\tau(s)}^{s_{|\beta|+1}} \cdots \int_{\tau(s)}^{s_{m-1}} \tilde{\mathbb{E}} [|L^{\alpha(\beta)} g^i(X_{s_m})|^2 | \mathcal{Y}_t] ds_m \cdots ds_{|\beta|+1} \\ & \leq \delta^{m-|\beta|} \int_{\tau(s)}^{s_{|\beta|}} \int_{\tau(s)}^{s_{|\beta|+1}} \cdots \int_{\tau(s)}^{s_{m-1}} \tilde{\mathbb{E}} [(1 + |X_{\tau(s)}|^r)^2] ds_m \cdots ds_{|\beta|+1} \\ & \leq C \delta^{m-|\beta|} \int_{\tau(s)}^{s_{|\beta|}} \int_{\tau(s)}^{s_{|\beta|+1}} \cdots \int_{\tau(s)}^{s_{m-1}} ds_m \cdots ds_{|\beta|+1}, \end{aligned}$$

and we can write

$$\begin{aligned} & \tilde{\mathbb{E}} \left[\hat{I}_{\beta}(\hat{\psi}_{i, \beta}^{\tau, m}(\cdot))_{0, s} \Theta_s^{g^i, \tau, m} | \mathcal{Y}_t \right]^2 \\ &= \tilde{\mathbb{E}} \left[I_{\alpha(\beta)}(L^{\alpha(\beta)} g^i(X))_{\tau(s), s} \hat{I}_{\beta}(\hat{\psi}_{i, \beta}^{\tau, m}(\cdot))_{\tau(s), s} | \mathcal{Y}_t \right]^2 \end{aligned}$$

$$\begin{aligned}
&= \tilde{\mathbb{E}} \left[\int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{|\beta|-1}} \Lambda_{\alpha(\beta)}(s_{|\beta|}) \hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|}) ds_{|\beta|} \cdots ds_1 | \mathcal{Y}_t \right]^2 \\
&\leq \delta^{|\beta|} \int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{|\beta|-1}} \tilde{\mathbb{E}} \left[\Lambda_{\alpha(\beta)}(s_{|\beta|}) \hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|}) | \mathcal{Y}_t \right]^2 ds_{|\beta|} \cdots ds_1 \\
&\leq \delta^{|\beta|} \int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{|\beta|-1}} \tilde{\mathbb{E}} \left[\Lambda_{\alpha(\beta)}(s_{|\beta|})^2 | \mathcal{Y}_t \right] \tilde{\mathbb{E}} \left[\hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|})^2 | \mathcal{Y}_t \right] ds_{|\beta|} \cdots ds_1 \\
&\leq C \delta^m \int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{|\beta|-1}} \int_{\tau(s)}^{s_{|\beta|}} \cdots \int_{\tau(s)}^{s_{m-1}} ds_m \cdots ds_{|\beta|+1} \\
&\quad \times \tilde{\mathbb{E}} \left[\hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|})^2 | \mathcal{Y}_t \right] ds_{|\beta|} \cdots ds_1 \\
&= C \delta^m \int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{m-1}} \tilde{\mathbb{E}} \left[\hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|})^2 | \mathcal{Y}_t \right] ds_m \cdots ds_1.
\end{aligned}$$

Finally,

$$\begin{aligned}
&\tilde{\mathbb{E}} \left[\tilde{\mathbb{E}} \left[(\xi_i - \xi_i^{\tau,m}) \hat{I}_\beta(\hat{\psi}_{i,\beta}^{\tau,m}(\cdot))_{0,t} | \mathcal{Y}_t \right]^{2p} \right] \\
&= \tilde{\mathbb{E}} \left[\left(\int_0^t \tilde{\mathbb{E}} [\hat{I}_\beta(\hat{\psi}_{i,\beta}^{\tau,m}(\cdot))_{0,s} \Theta_s^{g^i, \tau, m} | \mathcal{Y}_t] dY_s^i \right)^{2p} \right] \\
&\leq \tilde{\mathbb{E}} \left[\left(\int_0^t \tilde{\mathbb{E}} [\hat{I}_\beta(\hat{\psi}_{i,\beta}^{\tau,m}(\cdot))_{0,s} \Theta_s^{g^i, \tau, m} | \mathcal{Y}_t]^2 ds \right)^p \right] \\
&\leq C(p) \delta^{pm} \tilde{\mathbb{E}} \left[\left(\int_0^t \int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{m-1}} \tilde{\mathbb{E}} \left[\left| \hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|}) \right|^2 | \mathcal{Y}_t \right] ds_m \cdots ds_1 ds \right)^p \right] \\
&\leq C(p, t) \delta^{2pm-1} \int_0^t \int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{m-1}} \tilde{\mathbb{E}} \left[\left| \hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|}) \right|^{2p} \right] ds_m \cdots ds_1 ds \\
&\leq C(p, t) \delta^{2pm} \sup_{\tau \in \Pi} \sup_{0 \leq s_{|\beta|} < \cdots < s_1 \leq t} \tilde{\mathbb{E}} \left[\left| \hat{\psi}_{i,\beta}^{\tau,m}(s_1, \dots, s_{|\beta|}) \right|^{2p} \right] \\
&\leq C \delta^{2pm}.
\end{aligned}$$

Term $A_{i,3}$:

For any $i = 1, \dots, d_Y$, $\beta = (\beta_1, \dots, \beta_m) \in \mathcal{M}_{m-1}^R(S_1)$, we can use integration by parts to obtain

$$\begin{aligned}
(\xi_i - \xi_i^{\tau,m}) \hat{I}_\beta(\hat{\psi}_{i,\beta}^{\tau,m}(\cdot))_{0,t} &= \int_0^t \Theta_s^{g^i, \tau, m} dY_s^i \int_0^t \hat{I}_{\beta_-}(\hat{\psi}_{i,\beta}^{\tau,m}(\cdot, s))_{0,s} dV_s^{\beta_1} \\
&= \int_0^t \left(\int_0^s \Theta_u^{g^i, \tau, m} dY_u^i \right) \hat{I}_{\beta_-}(\hat{\psi}_{i,\beta}^{\tau,m}(\cdot, s))_{0,s} dV_s^{\beta_1} \\
&\quad + \int_0^t \hat{I}_\beta(\hat{\psi}_{i,\beta}^{\tau,m}(\cdot))_{0,s} \Theta_s^{g^i, \tau, m} dY_s^i.
\end{aligned}$$

As for the term $A_{i,2}$, one can show that

$$\int_0^v \left(\int_0^s \Theta_u^{g^i, \tau, m} dY_u^i \right) \hat{I}_{\beta-}(\psi_{i,\beta}^{\tau, m}(\cdot, s))_{0,s} dV_s^{\beta_1}, \quad v \in [0, t],$$

is a $\mathcal{F}_v \vee \mathcal{Y}_t$ -martingale and it vanishes when taking conditional expectation with respect to $\mathcal{F}_0 \vee \mathcal{Y}_t$. Therefore,

$$\begin{aligned} & \tilde{\mathbb{E}} \left[(\xi_i - \xi_i^{\tau, m}) \hat{I}_{\beta}(\psi_{i,\beta}^{\tau, m}(\cdot))_{0,t} | \mathcal{Y}_t \right] \\ &= \int_0^t \tilde{\mathbb{E}} \left[\hat{I}_{\beta}(\psi_{i,\beta}^{\tau, m}(\cdot))_{0,s} \Theta_s^{g^i, \tau, m} | \mathcal{Y}_t \right] dY_s^i \\ &= \int_0^t \tilde{\mathbb{E}} \left[\int_0^s \cdots \int_0^{s_{m-1}} \psi_{i,\beta}^{\tau, m}(s_1, \dots, s_m) dV_{s_1}^{\beta_1} \cdots dV_{s_m}^{\beta_m} \right. \\ &\quad \times \left. \left(\sum_{\alpha \in \mathcal{M}_{m-1}^R(S_0)} I_{\alpha}(L^{\alpha} g^i(X_{\cdot}))_{\tau(s), s} \right) | \mathcal{Y}_t \right] dY_s^i \\ &= \int_0^t \tilde{\mathbb{E}} \left[\int_0^s \cdots \int_0^{s_{m-1}} \psi_{i,\beta}^{\tau, m}(s_1, \dots, s_m) dV_{s_1}^{\beta_1} \cdots dV_{s_m}^{\beta_m} \right. \\ &\quad \times \left. \left(\sum_{\alpha \in \mathcal{M}_{m-1}^R(S_0)} \int_0^s \cdots \int_0^{s_2} \prod_{j=1}^m \mathbf{1}_{\{s_j > \tau(s)\}} L^{\alpha} g^i(X_{s_1}) dV_{s_1}^{\alpha_1} \cdots dV_{s_m}^{\alpha_m} \right) | \mathcal{Y}_t \right] dY_s^i, \end{aligned}$$

where $dV_{s_j}^{\alpha_j} = ds_j$ if $\alpha_j = 0$. Taking conditional expectation with respect to $\mathcal{F}_0 \vee \mathcal{Y}_t$, the only term that does not vanish is the one corresponding to $\alpha = \alpha(\beta) \triangleq (\beta_m, \dots, \beta_1)$ and we get that

$$\begin{aligned} & \tilde{\mathbb{E}} \left[\hat{I}_{\beta}(\psi_{i,\beta}^{\tau, m}(\cdot))_{0,s} \Theta_s^{g^i, \tau, m} | \mathcal{Y}_t \right]^2 \\ &= \tilde{\mathbb{E}} \left[\int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{m-1}} \tilde{\mathbb{E}} \left[\psi_{i,\beta}^{\tau, m}(s_1, \dots, s_m) L^{\alpha(\beta)} g^i(X_{s_m}) | \mathcal{F}_0 \vee \mathcal{Y}_t \right] ds_m \cdots ds_1 | \mathcal{Y}_t \right]^2 \\ &\leq \delta^m \tilde{\mathbb{E}} \left[\int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{m-1}} \tilde{\mathbb{E}} \left[\psi_{i,\beta}^{\tau, m}(s_1, \dots, s_m)^2 | \mathcal{F}_0 \vee \mathcal{Y}_t \right] \tilde{\mathbb{E}} \left[|L^{\beta} g^i(X_{s_1})|^2 | \mathcal{F}_0 \vee \mathcal{Y}_t \right] ds_m \cdots ds_1 | \mathcal{Y}_t \right] \\ &\leq C \delta^m \int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{m-1}} \tilde{\mathbb{E}} \left[(1 + |X_{\tau(s)}|^r)^2 \right] \tilde{\mathbb{E}} \left[\psi_{i,\beta}^{\tau, m}(s_1, \dots, s_m)^2 | \mathcal{Y}_t \right] ds_m \cdots ds_1 \\ &\leq C \delta^m \int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{m-1}} \tilde{\mathbb{E}} \left[\psi_{i,\beta}^{\tau, m}(s_1, \dots, s_m)^2 | \mathcal{Y}_t \right] ds_m \cdots ds_1. \end{aligned}$$

Therefore,

$$\begin{aligned} & \tilde{\mathbb{E}} \left[\tilde{\mathbb{E}} \left[(\xi_i - \xi_i^{\tau, m}) \hat{I}_{\beta}(\psi_{i,\beta}^{\tau, m}(\cdot))_{0,t} | \mathcal{Y}_t \right]^{2p} \right] \\ &= \tilde{\mathbb{E}} \left[\left(\int_0^t \tilde{\mathbb{E}} \left[\hat{I}_{\beta}(\psi_{i,\beta}^{\tau, m}(\cdot))_{0,s} \Theta_s^{g^i, \tau, m} | \mathcal{Y}_t \right] dY_s^i \right)^{2p} \right] \end{aligned}$$

$$\begin{aligned}
&\leq \tilde{\mathbb{E}} \left[\left(\int_0^t \tilde{\mathbb{E}}[\hat{I}_\beta(\psi_{i,\beta}^{\tau,m}(\cdot))_{0,s} \Theta_s^{g^i, \tau, m} | \mathcal{Y}_t]^2 ds \right)^p \right] \\
&\leq C(p) \delta^{pm} \tilde{\mathbb{E}} \left[\left(\int_0^t \int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{m-1}} \tilde{\mathbb{E}}[\psi_{i,\beta}^{\tau,m}(s_1, \dots, s_m)^2 | \mathcal{Y}_t] ds_m \cdots ds_1 ds \right)^p \right] \\
&\leq C(p, t) \delta^{2pm-1} \int_0^t \int_{\tau(s)}^s \cdots \int_{\tau(s)}^{s_{m-1}} \tilde{\mathbb{E}} \left[\left| \psi_{i,\beta}^{\tau,m}(s_1, \dots, s_m) \right|^{2p} \right] ds_m \cdots ds_1 ds \\
&\leq C(p, t) \delta^{2pm} \sup_{\tau \in \Pi} \sup_{0 \leq s_m < \cdots < s_1 \leq t} \tilde{\mathbb{E}} \left[\left| \psi_{i,\beta}^{\tau,m}(s_1, \dots, s_m) \right|^{2p} \right] \\
&\leq C \delta^{2pm}.
\end{aligned}$$

□

5. Proof of Theorem 1

We will deduce the result from Theorem 12. Therefore, we have to verify that conditions $\mathbf{L}(p, 2, \Pi(t, \delta_0))$ and $\mathbf{UK}(p, 2, \Pi(t, \delta_0))$ are satisfied for the particular setting of the filtering problem, where δ_0 is given by equation (3.2). Note that in this case the function $\psi : \mathbb{R}^{d_Y+1} \rightarrow \mathbb{R}$ is given by

$$\psi(z) = \mathbf{exp}(z) \triangleq \exp\left(\sum_{i=0}^{d_Y} z_i\right), \quad z \in \mathbb{R}^{d_Y+1},$$

and, for $i = 0, \dots, d_Y$, we have

$$\eta_i^{\tau,2} = \int_0^1 \varphi(X_t) \partial_i \mathbf{exp}(s\xi + (1-s)\xi^{\tau,2}) ds = \int_0^1 \varphi(X_t) \mathbf{exp}(s\xi + (1-s)\xi^{\tau,2}) ds,$$

where ξ and $\xi^{\tau,2}$ are computed with $g_i = h_i, i = 1, \dots, d_Y, g_0 = -\frac{1}{2}(h_1^2 + \cdots + h_{d_Y}^2)$.

Before we can proceed we require some preliminary results. The next two lemmas are needed to verify condition $\mathbf{L}(p, 2, \Pi(t, \delta_0))$.

Lemma 13. *Let $h \in \mathcal{B}_b$. Then, for any $p \in \mathbb{R}$ one has*

$$\tilde{\mathbb{E}}[Z_t^p] = \tilde{\mathbb{E}}[\mathbf{exp}(p\xi)] < \infty.$$

Proof. We have that

$$\begin{aligned}
\tilde{\mathbb{E}}[\mathbf{exp}(p\xi)] &= \tilde{\mathbb{E}} \left[\exp \left(p \sum_{i=1}^{d_Y} \int_0^t h^i(X_s) dY_s^i - \frac{p}{2} \sum_{i=1}^{d_Y} \int_0^t h^i(X_s)^2 ds \right) \right] \\
&= \tilde{\mathbb{E}} \left[\mathcal{E} \left(p \sum_{i=1}^{d_Y} \int_0^\cdot h^i(X_s) dY_s^i \right)_t \exp \left(\frac{p^2 - p}{2} \sum_{i=1}^{d_Y} \int_0^t h^i(X_s)^2 ds \right) \right]
\end{aligned}$$

$$\begin{aligned}
&\leq \exp\left(\frac{p^2 + |p|}{2} d_Y t \|h\|_\infty^2\right) \tilde{\mathbb{E}} \left[\mathcal{E} \left(p \sum_{i=1}^{d_Y} \int_0^\cdot h^i(X_s) dY_s^i \right)_t \right] \\
&= \exp\left(\frac{p^2 + |p|}{2} d_Y t \|h\|_\infty^2\right) < \infty,
\end{aligned}$$

where $\mathcal{E}(p \sum_{i=1}^{d_Y} \int_0^\cdot h^i(X_s) dY_s^i)_t$ denotes the stochastic exponential, which is a (genuine) martingale by Novikov's criterion with expectation equal to 1. \square

Lemma 14. Assume that $f, \sigma \in \mathcal{B}_b$ and $h \in \mathcal{B}_b \cap C_b^2$. Let $p \geq 1$ be fixed and τ be a partition with mesh size

$$\delta < \left(p \|Lh\|_\infty \sqrt{d_Y d_V} \right)^{-1},$$

where

$$\|Lh\|_\infty \triangleq \max_{\substack{i=1, \dots, d_Y \\ r=1, \dots, d_V}} \|L^r h^i\|_\infty.$$

Then, one has that

$$\tilde{\mathbb{E}} [\mathbf{exp}(p\xi^{\tau,2})] < \infty.$$

The proof of Lemma 14 is quite technical and is done in the last section. The next two lemmas are crucial to verify Condition $\mathbf{UK}(p, 2, \Pi(t, \delta_0))$.

Lemma 15. If $X_t \in \mathbb{R}^{d_X}$ is the solution to

$$X_t = x + \int_0^t A_0(X_s) ds + \sum_{j=1}^{d_V} \int_0^t A_j(X_s) dV_s^j,$$

where A_0, A_1, \dots, A_{d_V} are N -times continuously differentiable with bounded derivatives of order greater or equal than one and $V_t = (V_t^1, \dots, V_t^{d_V})$ is a d_V -dimensional Brownian motion. Then, $X_t^i \in \mathbb{D}^{N,p}$, $p \geq 1$, $t \in [0, T]$, $i = 1, \dots, d_X$. Furthermore, for any $p \geq 1$ one has that

$$\sup_{r_1, r_2, \dots, r_k \in [0, T]} \mathbb{E} \left[\sup_{r_1 \vee r_2 \vee \dots \vee r_k \leq t \leq T} |D_{r_1, r_2, \dots, r_k}^{j_1, j_2, \dots, j_k} X_t^i|^p \right] < \infty, \quad 1 \leq k \leq N.$$

Proof. See Nualart [8], Theorem 2.2.1. and 2.2.2. \square

Lemma 16. Assume $f, \sigma \in \mathcal{B}_b \cap C_b^2$, $h \in \mathcal{B}_b \cap C_b^4$ and $\varphi \in C_P^2$. Then, the random vector $\eta^{\tau,2} = (\eta_i^{\tau,2})_{i=0}^{d_Y}$ belongs to $\mathbb{D}^{2,p}$, $p \geq 1$. Moreover,

$$\sup_{r_1, \dots, r_{|\alpha|} \in [0, t]} \tilde{\mathbb{E}} \left[\left| D_{r_1, \dots, r_{|\alpha|}}^{\alpha_1, \dots, \alpha_{|\alpha|}} \eta_i^{\tau,2} \right|^p \right] < \infty,$$

for any $i \in \{0, \dots, d_Y\}$ and $\alpha \in \mathcal{M}_2(S_1)$.

The proof of Lemma 16 is done in the last section.

Remark 17. The proof of Lemma 16 can be adapted to the case of $\psi \in C_P^{m+1}(\mathbb{R}^{d_Y+1}; \mathbb{R})$ without any requirement on the partition mesh. Hence, if we assume that $\psi \in C_P^{m+1}$, $\varphi \in C_P^m$, $f, \sigma \in C_b^m$ and that condition $\mathbf{G}(m)$ holds, then conditions $\mathbf{L}(p, m, \Pi(t))$ and $\mathbf{UK}(p, m, \Pi(t))$ also hold and Theorem 12 can be applied.

We are finally ready to put everything together and deduce Theorem 1.

Proof of Theorem 1. We will deduce the result from Theorem 12. Hence, we only need to check that conditions $\mathbf{S}(2), \mathbf{G}(2), \mathbf{L}(p, 2, \Pi(t, \delta_0))$ and $\mathbf{UK}(p, 2, \Pi(t, \delta_0))$ are satisfied. As $f, \sigma \in C_b^2$ and X_0 has moments of all orders, condition $\mathbf{S}(2)$ is satisfied. Moreover, as $g_i = h_i, i = 1, \dots, d_Y, g_0 = -\frac{1}{2}(h_1^2 + \dots + h_{d_Y}^2)$ and $h \in C_P^4$ we also have that condition $\mathbf{G}(2)$ is satisfied. By Hölder inequality and inequality (6.4) we get that

$$\begin{aligned} & \tilde{\mathbb{E}} \left[|\varphi(X_t) \partial_i \psi(s\xi + (1-s)\xi^{\tau,2})|^{2p+\varepsilon} \right] \\ & \leq \tilde{\mathbb{E}} \left[|\varphi(X_t)|^{2p+\varepsilon} \mathbf{exp}((2p+\varepsilon)\{s\xi + (1-s)\xi^{\tau,2}\}) \right] \\ & \leq \tilde{\mathbb{E}} \left[|\varphi(X_t)|^{\frac{(2p+\varepsilon)(2p+\varepsilon')}{\varepsilon'-\varepsilon}} \right]^{(\varepsilon'-\varepsilon)/(2p+\varepsilon')} \\ & \quad \times \tilde{\mathbb{E}} \left[\mathbf{exp}((2p+\varepsilon')\{s\xi + (1-s)\xi^{\tau,2}\}) \right]^{(2p+\varepsilon)/(2p+\varepsilon')} \\ & \leq C(p, \varepsilon, \varepsilon') \left(\tilde{\mathbb{E}} [\mathbf{exp}((2p+\varepsilon')\xi)] + \tilde{\mathbb{E}} [\mathbf{exp}((2p+\varepsilon')\xi^{\tau,2})] \right)^{(2p+\varepsilon)/(2p+\varepsilon')}, \end{aligned}$$

where $\varepsilon' > \varepsilon > 0$ are such that $\tilde{\mathbb{E}} [\mathbf{exp}((2p+\varepsilon')\xi^{\tau,2})] < \infty$, which exist due to Lemma 14 and the fact that $\delta < \delta_0$. Note that we can apply Lemma 14 because $f, \sigma \in \mathcal{B}_b$ and $h \in \mathcal{B}_b \cap C_b^2$. Combining with Lemma 13 we can conclude that condition $\mathbf{L}(p, 2, \Pi(t, \delta_0))$ holds. Moreover, condition $\mathbf{UK}(p, 2, \Pi(t, \delta_0))$ holds due to Lemma 16 and Theorem 9. Note that we can apply Lemma 16 because $f, \sigma \in \mathcal{B}_b \cap C_b^2, h \in \mathcal{B}_b \cap C_b^4$ and $\varphi \in C_P^2$. Next, applying Theorem 12 we get the desired rate of convergence for the unnormalised conditional distribution $\rho_t^{\tau,2}$. To prove the rate for the normalised conditional distribution observe that we can write

$$\pi_t^{\tau,2}(\varphi) - \pi_t(\varphi) = \frac{1}{\rho_t(\mathbf{1})} \frac{\rho_t^{\tau,2}(\varphi)}{\rho_t^{\tau,2}(\mathbf{1})} \left(\rho_t(\mathbf{1}) - \rho_t^{\tau,2}(\mathbf{1}) \right) + \frac{1}{\rho_t(\mathbf{1})} \left(\rho_t^{\tau,2}(\varphi) - \rho_t(\varphi) \right),$$

hence

$$\begin{aligned} & \mathbb{E} \left[\left| \pi_t(\varphi) - \pi_t^{\tau,2}(\varphi) \right|^p \right] \\ & \leq C(p) \tilde{\mathbb{E}} \left[\frac{Z_t}{|\rho_t(\mathbf{1})|^p} \left\{ \left| \pi_t^{\tau,2}(\varphi) \right|^p \left| \rho_t(\mathbf{1}) - \rho_t^{\tau,2}(\mathbf{1}) \right|^p + \left| \rho_t^{\tau,2}(\varphi) - \rho_t(\varphi) \right|^p \right\} \right] \\ & = C(p) \tilde{\mathbb{E}} \left[\frac{\tilde{\mathbb{E}}[Z_t|\mathcal{Y}_t]}{|\rho_t(\mathbf{1})|^p} \left\{ \left| \pi_t^{\tau,2}(\varphi) \right|^p \left| \rho_t(\mathbf{1}) - \rho_t^{\tau,2}(\mathbf{1}) \right|^p + \left| \rho_t^{\tau,2}(\varphi) - \rho_t(\varphi) \right|^p \right\} \right] \\ & \leq C(p) \left\{ \tilde{\mathbb{E}} \left[|\rho_t(\mathbf{1})|^{(1-p)} \left| \pi_t^{\tau,2}(\varphi) \right|^p \left| \rho_t(\mathbf{1}) - \rho_t^{\tau,2}(\mathbf{1}) \right|^p \right] \right. \end{aligned}$$

$$+ \tilde{\mathbb{E}} \left[|\rho_t(\mathbf{1})|^{(1-p)} \left| \rho_t^{\tau,2}(\varphi) - \rho_t(\varphi) \right|^p \right] \Big\} \\ \triangleq C(p) \{A_1 + A_2\}.$$

Applying Hölder inequality, we obtain

$$\begin{aligned} A_1 &\leq \tilde{\mathbb{E}} \left[|\rho_t(\mathbf{1})|^{2(1-p)} \left| \pi_t^{\tau,2}(\varphi) \right|^{2p} \right]^{1/2} \tilde{\mathbb{E}} \left[\left| \rho_t(\mathbf{1}) - \rho_t^{\tau,2}(\mathbf{1}) \right|^{2p} \right]^{1/2} \\ &\leq \tilde{\mathbb{E}} \left[|\rho_t(\mathbf{1})|^{2(1-p)(2p+\varepsilon)/\varepsilon} \right]^{\varepsilon/(2(2p+\varepsilon))} \tilde{\mathbb{E}} \left[\left| \pi_t^{\tau,2}(\varphi) \right|^{2p+\varepsilon} \right]^{2p/(2(2p+\varepsilon))} \\ &\quad \times \tilde{\mathbb{E}} \left[\left| \rho_t(\mathbf{1}) - \rho_t^{\tau,2}(\mathbf{1}) \right|^{2p} \right]^{1/2}, \end{aligned}$$

and

$$A_2 \leq \tilde{\mathbb{E}} \left[|\rho_t(\mathbf{1})|^{2(1-p)} \right]^{1/2} \tilde{\mathbb{E}} \left[\left| \rho_t^{\tau,2}(\varphi) - \rho_t(\varphi) \right|^{2p} \right]^{1/2}.$$

Combining the bounds for the unnormalised distribution, the hypothesis on $\pi_t^{\tau,2}(\varphi)$ and the fact that, due to Lemma 13, for any $q \leq 0$ we have that

$$\tilde{\mathbb{E}} [|\rho_t(\mathbf{1})|^q] = \tilde{\mathbb{E}} \left[\left| \tilde{\mathbb{E}}[Z_t | \mathcal{Y}_t] \right|^q \right] \leq \tilde{\mathbb{E}} [Z_t^q] < \infty,$$

we can conclude. \square

Remark 18. The assumption $\sup_{\tau \in \Pi(t, \delta_0)} \tilde{\mathbb{E}} \left[\left| \pi_t^{\tau,2}(\varphi) \right|^{2p+\varepsilon} \right] < \infty$ for some $\varepsilon > 0$ is satisfied if φ is bounded. If φ is unbounded, note that by using Jensen's inequality one has

$$\begin{aligned} \tilde{\mathbb{E}} \left[\left| \pi_t^{\tau,2}(\varphi) \right|^{2p+\varepsilon} \right] &= \tilde{\mathbb{E}} \left[\left| \tilde{\mathbb{E}} \left[\frac{\varphi(X_t) Z_t^{\tau,2}}{\tilde{\mathbb{E}}[Z_t^{\tau,2} | \mathcal{Y}_t]} \middle| \mathcal{Y}_t \right] \right|^{2p+\varepsilon} \right] \\ &\leq \tilde{\mathbb{E}} \left[|\varphi(X_t)|^{2p+\varepsilon} \exp((2p+\varepsilon)(\xi^{\tau,2} - \tilde{\mathbb{E}}[\xi^{\tau,2} | \mathcal{Y}_t])) \right]. \end{aligned}$$

Hence, if φ has polynomial growth and $h \in \mathcal{B}_b \cap C_b^2$, one can reason as in Lemma 14 to obtain $\sup_{\tau \in \Pi(t, \delta_0)} \tilde{\mathbb{E}} \left[\left| \pi_t^{\tau,2}(\varphi) \right|^{2p+\varepsilon} \right] < \infty$.

6. Proof of technical results

In this section we provide the proof for Lemmas 14 and 16, which are of a more technical nature.

Proof of Lemma 14. We can write $\exp(p\xi^{\tau,2}) \triangleq \prod_{i=1}^4 \left(K_t^{\tau,2,i} \right)^p$, where

$$K_t^{\tau,2,1} \triangleq \exp \left(\sum_{i=1}^{d_Y} \sum_{r=1}^{d_V} \int_0^t L^r h^i(X_{\tau(s)}) (V_s^r - V_{\tau(s)}^r) dY_s^i \right),$$

$$\begin{aligned}
K_t^{\tau,2,2} &\triangleq \exp \left(\sum_{i=1}^{d_Y} \int_0^t \{h^i(X_{\tau(s)}) + L^0 h^i(X_{\tau(s)})(s - \tau(s))\} dY_s^i \right), \\
K_t^{\tau,2,3} &\triangleq \exp \left(-\frac{1}{2} \sum_{i=1}^{d_Y} \int_0^t \{(h^i)^2(X_{\tau(s)}) + L^0((h^i)^2)(X_{\tau(s)})(s - \tau(s))\} ds \right), \\
K_t^{\tau,2,4} &\triangleq \exp \left(-\frac{1}{2} \sum_{i=1}^{d_Y} \sum_{r=1}^{d_V} \int_0^t L^r((h^i)^2)(X_{\tau(s)})(V_s^r - V_{\tau(s)}^r) ds \right).
\end{aligned}$$

Let $\varepsilon > 0$, then, by Hölder inequality, we have

$$\tilde{\mathbb{E}} [\mathbf{exp}(p\xi^{\tau,2})] \leq \tilde{\mathbb{E}} \left[\left| K_t^{\tau,2,1} \right|^{p(1+\varepsilon)} \right]^{\frac{1}{1+\varepsilon}} \tilde{\mathbb{E}} \left[\prod_{i=2}^4 \left| K_t^{\tau,2,i} \right|^{p \frac{(1+\varepsilon)}{\varepsilon}} \right]^{\frac{\varepsilon}{1+\varepsilon}}.$$

Hence, the result follows by showing that $K_t^{\tau,2,1}$ has finite $p(1+\varepsilon)$ -moment and

$$\tilde{\mathbb{E}} \left[\prod_{i=2}^4 \left| K_t^{\tau,2,i} \right|^{p \frac{(1+\varepsilon)}{\varepsilon}} \right] < \infty. \quad (6.1)$$

Applying Hölder inequality twice, condition (6.1) follows by showing that $K_t^{\tau,2,i}$, $i = 2, \dots, 4$ have finite moments of all orders. In what follows, let $q \geq 1$ be a fixed real constant. We start by the easiest term, $K_t^{\tau,2,3}$. We have that

$$\tilde{\mathbb{E}} \left[\left| K_t^{\tau,2,3} \right|^q \right] \leq \exp \left(\frac{qd_Y}{2} t (\|h\|_\infty^2 + \delta \|L^0 h^2\|_\infty) \right) < \infty,$$

because $\|h\|_\infty^2$ and $\|L^0 h^2\|_\infty = \max_{i=1, \dots, d_Y} \|L^0(h_i^2)\|_\infty$ are finite due to the assumptions on f, σ and h . For the term $K_t^{\tau,2,4}$, we can write

$$\begin{aligned}
\tilde{\mathbb{E}} \left[\left| K_t^{\tau,2,4} \right|^q \right] &\leq \tilde{\mathbb{E}} \left[\exp \left(\frac{qd_Y d_V}{2} \|L((h)^2)\|_\infty \int_0^t |V_s^1 - V_{\tau(s)}^1| ds \right) \right] \\
&\leq \tilde{\mathbb{E}} \left[\exp \left(\frac{qd_Y d_V}{2} \|L((h)^2)\|_\infty t \sqrt{\delta} |V_1^1| \right) \right] < \infty,
\end{aligned}$$

because $\|L((h)^2)\|_\infty = \max_{i=1, \dots, d_Y} \max_{r=1, \dots, d_V} \|L^r(h_i^2)\|_\infty$ is finite and $|V_1^1|$ has exponential moments of any order.

For the term $K_t^{\tau,2,2}$, we first condition with respect to $\mathcal{F}_t^V = \sigma(V_s, 0 \leq s \leq t)$ and use the fact that, conditionally to \mathcal{F}_t^V , the stochastic integrals with respect to Y are Gaussian. We get

$$\begin{aligned}
&\tilde{\mathbb{E}} \left[\left| K_t^{\tau,2,2} \right|^q \right] \\
&= \tilde{\mathbb{E}} \left[\tilde{\mathbb{E}} \left[\exp \left(q \sum_{i=1}^{d_Y} \int_0^t \{h^i(X_{\tau(s)}) + L^0 h^i(X_{\tau(s)})(s - \tau(s))\} dY_s^i \right) \middle| \mathcal{F}_t^V \right] \right]
\end{aligned}$$

$$\begin{aligned}
&= \tilde{\mathbb{E}} \left[\exp \left(\frac{q^2}{2} \sum_{i=1}^{d_Y} \int_0^t \{h^i(X_{\tau(s)}) + L^0 h^i(X_{\tau(s)})(s - \tau(s))\}^2 ds \right) \right] \\
&= \exp(q^2 d_Y t \{\|h\|_\infty^2 + \|L^0 h\|_\infty^2\}) < \infty.
\end{aligned}$$

Finally, the term $K_t^{\tau,2,1}$ is more delicate because, in order to show that has finite $(p + \varepsilon)$ -moment, a relationship between the mesh of the partition δ and $p + \varepsilon$ is needed. Proceeding as with the term $K_t^{\tau,2,2}$, we obtain

$$\begin{aligned}
&\tilde{\mathbb{E}} \left[\left| K_t^{\tau,2,1} \right|^{p(1+\varepsilon)} \right] \\
&= \tilde{\mathbb{E}} \left[\exp \left(p(1+\varepsilon) \sum_{i=1}^{d_Y} \sum_{r=1}^{d_V} \int_0^t L^r h^i(X_{\tau(s)})(V_s^r - V_{\tau(s)}^r) dY_s^i \right) \right] \\
&= \tilde{\mathbb{E}} \left[\prod_{i=1}^{d_Y} \tilde{\mathbb{E}} \left[\exp \left(\int_0^t p(1+\varepsilon) \sum_{r=1}^{d_V} L^r h^i(X_{\tau(s)})(V_s^r - V_{\tau(s)}^r) dY_s^i \right) | \mathcal{F}_t^V \right] \right].
\end{aligned}$$

Now, conditionally to \mathcal{F}_t^V , the terms in the exponential are centered Gaussian random variables and we get that

$$\begin{aligned}
&\tilde{\mathbb{E}} \left[\left| K_t^{\tau,2,1} \right|^{p(1+\varepsilon)} \right] \\
&= \tilde{\mathbb{E}} \left[\prod_{i=1}^{d_Y} \exp \left(\frac{p^2(1+\varepsilon)^2}{2} \int_0^t \left(\sum_{r=1}^{d_V} L^r h^i(X_{\tau(s)})(V_s^r - V_{\tau(s)}^r) \right)^2 ds \right) \right] \\
&\leq \tilde{\mathbb{E}} \left[\prod_{i=1}^{d_Y} \exp \left(\frac{p^2(1+\varepsilon)^2 d_V}{2} \int_0^t \left(\sum_{r=1}^{d_V} |L^r h^i(X_{\tau(s)})|^2 (V_s^r - V_{\tau(s)}^r)^2 \right) ds \right) \right] \\
&= \tilde{\mathbb{E}} \left[\exp \left(\frac{p^2(1+\varepsilon)^2 d_Y d_V \|Lh\|_\infty^2}{2} \sum_{r=1}^{d_V} \int_0^t (V_s^r - V_{\tau(s)}^r)^2 ds \right) \right] \\
&= \tilde{\mathbb{E}} \left[\exp \left(\frac{p^2(1+\varepsilon)^2 d_Y d_V \|Lh\|_\infty^2}{2} \int_0^t (V_s^1 - V_{\tau(s)}^1)^2 ds \right) \right]^{d_V}.
\end{aligned}$$

So we need to find conditions on $\beta > 0$, such that $\tilde{\mathbb{E}} \left[\exp \left(\beta \int_0^t (V_s^1 - V_{\tau(s)}^1)^2 ds \right) \right] < \infty$. We can write

$$\begin{aligned}
\tilde{\mathbb{E}} \left[\exp \left(\beta \int_0^t (V_s^1 - V_{\tau(s)}^1)^2 ds \right) \right] &= \tilde{\mathbb{E}} \left[\exp \left(\beta \sum_{j=1}^n \int_{t_{j-1}}^{t_j} (V_s^1 - V_{t_{j-1}}^1)^2 ds \right) \right] \\
&= \prod_{j=1}^n \tilde{\mathbb{E}} \left[\exp \left(\beta \int_{t_{j-1}}^{t_j} (V_s^1 - V_{t_{j-1}}^1)^2 ds \right) \right] \\
&\triangleq \prod_{j=1}^n \Theta(\beta, \delta_j).
\end{aligned}$$

Denote by $M_t \triangleq \sup_{0 \leq s \leq t} V_s^1$ and recall that the density of M_t is given by $f_{M_t}(x) = \frac{2}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \mathbf{1}_{(0,\infty)}$. Moreover, note that for any $A > 0$,

$$\frac{2}{\sqrt{2\pi\sigma^2}} \int_0^\infty \exp\left\{-A \frac{x^2}{2\sigma^2}\right\} dx = A^{-1/2}.$$

Then, we have that

$$\begin{aligned} \Theta(\beta, \delta_j) &\leq \tilde{\mathbb{E}}[\exp(\beta \delta_j M_\delta^2)] = \int_0^\infty \frac{2}{\sqrt{2\pi\delta_j}} \exp\left\{\beta \delta_j x^2 - \frac{x^2}{2\delta_j}\right\} \\ &= \int_0^\infty \frac{2}{\sqrt{2\pi\delta_j}} \exp\left\{-\left(1 - 2\beta\delta_j^2\right) \frac{x^2}{2\delta_j}\right\} = (1 - 2\beta\delta_j^2)^{-1/2} < \infty, \end{aligned}$$

as long as $1 - 2\beta\delta_j^2 > 0$. On the other hand,

$$\begin{aligned} (1 - 2\beta\delta_j^2)^{-1} &= \sum_{k=0}^\infty (2\beta\delta_j^2)^k = 1 + 2\beta\delta_j^2 \left(\sum_{k=0}^\infty (2\beta\delta_j^2)^k \right) \\ &\leq 1 + 2\beta\delta_j^2 \left(\sum_{k=0}^\infty (2\beta\delta^2)^k \right) = 1 + \frac{2\beta\delta_j^2}{1 - 2\beta\delta^2} \\ &\leq \exp\left(\frac{2\beta\delta_j^2}{1 - 2\beta\delta^2}\right), \end{aligned}$$

and, therefore,

$$\begin{aligned} \prod_{j=1}^n \Theta(\beta, \delta_j) &\leq \prod_{j=1}^n \exp\left(\frac{\beta\delta_j^2}{1 - 2\beta\delta^2}\right) \leq \exp\left(\frac{\beta \sum_{j=1}^n \delta_j^2}{1 - 2\beta\delta^2}\right) \\ &\leq \exp\left(\frac{\beta\delta t}{1 - 2\beta\delta^2}\right) < \infty. \end{aligned}$$

As $\beta = \frac{p^2(1+\varepsilon)^2 d_Y d_V \|Lh\|_\infty^2}{2}$ and $\varepsilon > 0$ can be made arbitrary small we get the following condition for the partition mesh $\delta < (p \|Lh\|_\infty \sqrt{d_Y d_V})^{-1}$. \square

Proof of Lemma 16. To ease the notation we are just going to give the proof for $d_V = d_Y = d_X = 1$. Let $F \triangleq \varphi(X_t)$ and $G \triangleq \int_0^1 \mathbf{exp}(s\xi + (1-s)\bar{\xi}^{\tau,2}) ds$. Then, by Leibniz's rule, for any $\alpha \in \mathcal{M}_2(S_1)$ we get

$$\begin{aligned} D_{r_1, \dots, r_{|\alpha|}}^{\alpha_1, \dots, \alpha_{|\alpha|}} \eta_i^{\tau,2} &= D_{r_1, \dots, r_{|\alpha|}}^{|\alpha|} \eta_i^{\tau,2} = D_{r_1, \dots, r_{|\alpha|}}^{|\alpha|} (FG) \\ &= \sum_{k=0}^{|\alpha|} \binom{|\alpha|}{k} (D_{r_1, \dots, r_k}^k F) (D_{r_1, \dots, r_{|\alpha|-k}}^{|\alpha|-k} G), \end{aligned}$$

and applying Hölder's inequality one has that

$$\tilde{\mathbb{E}} \left[\left| D_{r_1, \dots, r_{|\alpha|}}^{|\alpha|} \eta_i^{\tau,2} \right|^p \right]$$

$$\begin{aligned}
&\leq C(|\alpha|) \sum_{k=0}^{|\alpha|} \binom{|\alpha|}{k} \tilde{\mathbb{E}} \left[\left| (D_{r_1, \dots, r_k}^k F) (D_{r_1, \dots, r_{|\alpha|-k}}^{|\alpha|-k} G) \right|^p \right] \\
&\leq C(|\alpha|) \sum_{k=0}^{|\alpha|} \binom{|\alpha|}{k} \tilde{\mathbb{E}} \left[\left| D_{r_1, \dots, r_k}^k F \right|^{p \frac{(1+\varepsilon)}{\varepsilon}} \right]^{\varepsilon/(1+\varepsilon)} \\
&\quad \times \tilde{\mathbb{E}} \left[\left| D_{r_1, \dots, r_{|\alpha|-k}}^{|\alpha|-k} G \right|^{p(1+\varepsilon)} \right]^{1/(1+\varepsilon)},
\end{aligned}$$

for some $\varepsilon > 0$. Hence, the result follows if we show that

$$\sup_{r_1, \dots, r_k \in [0, t]} \tilde{\mathbb{E}} [|D_{r_1, \dots, r_k}^k F|^q] < \infty, \quad 0 \leq k \leq |\alpha|, \quad (6.2)$$

$$\sup_{r_1, \dots, r_k \in [0, t]} \tilde{\mathbb{E}} [|D_{r_1, \dots, r_k}^k G|^{p(1+\varepsilon)}] < \infty, \quad 0 \leq k \leq |\alpha|, \quad (6.3)$$

for any $q \geq 1$ and some $\varepsilon > 0$.

Proof of (6.2):

If $k = 0$, using that $f \in C_b^2, \sigma \in C_b^2$ and $\varphi \in C_P^2$, we have that $\tilde{\mathbb{E}}[|F|^q] = \tilde{\mathbb{E}}[|\varphi(X_t)|^q] < \infty$. If $1 \leq k \leq |\alpha|$, we use Faà di Bruno's formula to obtain an expression for $D_{r_1, \dots, r_k}^k F$ in terms of the so called partial Bell polynomials, which are given by

$$B_{k,a}(x_1, \dots, x_k) = \sum_{(j_1, \dots, j_k) \in \Lambda(k,a)} \frac{k!}{j_1! (1!)^{j_1} j_2! (2!)^{j_2} \dots j_k! (k!)^{j_k}} x_1^{j_1} x_2^{j_2} \dots x_k^{j_k},$$

where $1 \leq a \leq k$ and

$$\Lambda(k, a) = \{(j_1, \dots, j_k) \in \mathbb{Z}_+^k : j_1 + 2j_2 + \dots + kj_k = k, j_1 + j_2 + \dots + j_k = a\}.$$

In particular, we have that

$$D_{r_1, \dots, r_k}^k \varphi(X_t) = \sum_{a=1}^k \varphi^{(a)}(X_t) B_{k,a}(D_{r_1}^1 X_t, D_{r_1, r_2}^2 X_t, \dots, D_{r_1, \dots, r_k}^k X_t).$$

Hence, for any $q \geq 1$, applying Cauchy-Schwarz inequality we get

$$\begin{aligned}
\tilde{\mathbb{E}} [|D_{r_1, \dots, r_k}^k F|^q] &\leq C(q, k) \sum_{a=1}^k \tilde{\mathbb{E}} [|\varphi^{(a)}(X_t) B_{k,a}(D_{r_1}^1 X_t, D_{r_1, r_2}^2 X_t, \dots, D_{r_1, \dots, r_k}^k X_t)|^q] \\
&\leq C(q, k) \sum_{a=1}^k \tilde{\mathbb{E}} [|\varphi^{(a)}(X_t)|^{2q}]^{1/2} \\
&\quad \times \tilde{\mathbb{E}} [B_{k,a}(D_{r_1}^1 X_t, D_{r_1, r_2}^2 X_t, \dots, D_{r_1, \dots, r_k}^k X_t)^{2q}]^{1/2}.
\end{aligned}$$

The terms $\tilde{\mathbb{E}} [|\varphi^{(a)}(X_t)|^{2q}] < \infty, a = 1, \dots, k$, due to Remark 4 and that $\varphi \in C_P^2$. On the other hand, using a generalized version of Hölder's inequality we can bound

$$\tilde{\mathbb{E}} [B_{k,a}(D_{r_1}^1 X_t, D_{r_1, r_2}^2 X_t, \dots, D_{r_1, \dots, r_k}^k X_t)^{2q}], \quad 1 \leq a \leq k,$$

by a sum of products of expectations of powers of Malliavin derivatives of different orders. Combining this bound with Lemma 15 we get that the integrability condition (6.2) is satisfied.

Proof of (6.3):

First note that, by the convexity of the exponential function, we have that

$$\begin{aligned} \mathbf{exp}(q\{s\xi + (1-s)\xi^{\tau,2}\}) &= \exp\left(sq \sum_{i=0}^{d_Y} \xi_i + (1-s)q \sum_{i=0}^{d_Y} \xi_i^{\tau,2}\right) \\ &\leq s \exp\left(q \sum_{i=0}^{d_Y} \xi_i\right) + (1-s) \exp\left(q \sum_{i=0}^{d_Y} \xi_i^{\tau,2}\right) \\ &\leq \mathbf{exp}(q\xi) + \mathbf{exp}(q\xi_i^{\tau,2}), \end{aligned} \quad (6.4)$$

where $p > 0$ and $0 \leq s \leq 1$.

If $k = 0$, we have that

$$\begin{aligned} \tilde{\mathbb{E}}\left[|G|^{p(1+\varepsilon)}\right] &= \tilde{\mathbb{E}}\left[\left|\int_0^1 \mathbf{exp}(s\xi + (1-s)\xi^{\tau,2})ds\right|^{p(1+\varepsilon)}\right] \\ &\leq \int_0^1 \tilde{\mathbb{E}}[\mathbf{exp}(p(1+\varepsilon)(s\xi + (1-s)\xi^{\tau,2}))]ds \\ &\leq \tilde{\mathbb{E}}[\mathbf{exp}(p(1+\varepsilon)\xi)] + \tilde{\mathbb{E}}[\mathbf{exp}(p(1+\varepsilon)\xi^{\tau,2})] < \infty \end{aligned}$$

where we have used (6.4) and Lemmas 13 and 14. If $1 \leq k \leq |\alpha|$, using the basic properties of the Malliavin derivative and the definition of \mathbf{exp} , we have that

$$\begin{aligned} D_{r_1, \dots, r_k}^k G &= \int_0^1 D_{r_1, \dots, r_k}^k \mathbf{exp}(s\xi + (1-s)\xi^{\tau,2}) ds, \\ &= \int_0^1 D_{r_1, \dots, r_k}^k \exp\left(\sum_{i=0}^{d_Y} s\xi_i + (1-s)\xi_i^{\tau,2}\right) ds \\ &= \int_0^1 D_{r_1, \dots, r_k}^k \exp(\Theta_s) ds, \end{aligned} \quad (6.5)$$

where $\Theta_s \triangleq \sum_{i=0}^{d_Y} s\xi_i + (1-s)\xi_i^{\tau,2}$. Using again Faà di Bruno's formula we get

$$\begin{aligned} D_{r_1, \dots, r_k}^k \exp(\Theta_s) &= \sum_{a=1}^k \frac{d^a}{dx^a} \exp(\Theta_s) B_{k,a}(D_{r_1}^1 \Theta_s, D_{r_1, r_2}^2 \Theta_s, \dots, D_{r_1, \dots, r_k}^k \Theta_s) \\ &= \exp(\Theta_s) \sum_{a=1}^k B_{k,a}(D_{r_1}^1 \Theta_s, D_{r_1, r_2}^2 \Theta_s, \dots, D_{r_1, \dots, r_k}^k \Theta_s). \end{aligned} \quad (6.6)$$

and, on the other hand,

$$|B_{k,a}(D_{r_1}^1 \Theta_s, D_{r_1, r_2}^2 \Theta_s, \dots, D_{r_1, \dots, r_k}^k \Theta_s)|^{p(1+\varepsilon)}$$

$$\begin{aligned}
&\leq C(p, \varepsilon, k, a) \sum_{(j_1, \dots, j_k) \in \Lambda(k, a)} \left(\frac{k!}{j_1! (1!)^{j_1} \dots j_k! (k!)^{j_k}} \right)^{p(1+\varepsilon)} \\
&\times |D_{r_1}^1 \Theta_s|^{p(1+\varepsilon)j_1} \dots |D_{r_1, \dots, r_k}^k \Theta_s|^{p(1+\varepsilon)j_k} \\
&\triangleq \tilde{B}_{k,a}^{p,\varepsilon}(|D_{r_1}^1 \Theta_s|, |D_{r_1, r_2}^2 \Theta_s|, \dots, |D_{r_1, \dots, r_k}^k \Theta_s|). \tag{6.7}
\end{aligned}$$

Note also that by the linearity of the Malliavin derivative we have that

$$\begin{aligned}
|D_{r_1, \dots, r_a}^a \Theta_s| &= \left| D_{r_1, \dots, r_a}^a \left(\sum_{i=0}^{d_Y} s \xi_i + (1-s) \xi_i^{\tau,2} \right) \right| \\
&= \left| \sum_{i=0}^{d_Y} s D_{r_1, \dots, r_a}^a \xi_i + (1-s) D_{r_1, \dots, r_a}^a \xi_i^{\tau,2} \right| \\
&\leq \sum_{i=0}^{d_Y} |D_{r_1, \dots, r_a}^a \xi_i| + |D_{r_1, \dots, r_a}^a \xi_i^{\tau,2}|, \tag{6.8}
\end{aligned}$$

$1 \leq a \leq k$. Hence, combining equations (6.5), (6.6), (6.7), (6.8) and using Cauchy-Schwarz inequality we get

$$\begin{aligned}
&\tilde{\mathbb{E}}[|D_{r_1, \dots, r_k}^k G|^{p(1+\varepsilon)}] \\
&\leq \int_0^1 \tilde{\mathbb{E}}[|D_{r_1, \dots, r_k}^k \exp(\Theta_s)|^{p(1+\varepsilon)}] ds \\
&\leq \tilde{\mathbb{E}}[\mathbf{exp}(p(1+\varepsilon)\xi) \Phi(\xi, \xi^{\tau,2})] + \tilde{\mathbb{E}}[\mathbf{exp}(p(1+\varepsilon)\xi^{\tau,2}) \Phi(\xi^{\tau,2})] \\
&\leq \left\{ \tilde{\mathbb{E}}[\mathbf{exp}(p(1+\varepsilon')\xi)]^{(1+\varepsilon)/(1+\varepsilon')} + \tilde{\mathbb{E}}[\mathbf{exp}(p(1+\varepsilon')\xi^{\tau,2})]^{(1+\varepsilon)/(1+\varepsilon')} \right\} \\
&\times \tilde{\mathbb{E}} \left[|\Phi(\xi, \xi^{\tau,2})|^{\frac{1+\varepsilon'}{\varepsilon'-\varepsilon}} \right]^{(\varepsilon'-\varepsilon)/(1+\varepsilon')},
\end{aligned}$$

where $\varepsilon' > \varepsilon$ and

$$\begin{aligned}
\Phi(\xi, \xi^{\tau,2}) &\triangleq C(k, p) \sum_{a=1}^k C(p, k, a) \sum_{(j_1, \dots, j_k) \in \Lambda(k, a)} \left(\frac{k!}{j_1! (1!)^{j_1} \dots j_k! (k!)^{j_k}} \right)^p \\
&\times \left| \sum_{i=0}^{d_Y} |D_{r_1}^1 \xi_i| + |D_{r_1}^1 \xi_i^{\tau,2}| \right|^{pj_1} \dots \left| \sum_{i=0}^{d_Y} |D_{r_1, \dots, r_k}^k \xi_i| + |D_{r_1, \dots, r_k}^k \xi_i^{\tau,2}| \right|^{pj_k}.
\end{aligned}$$

The integrability of $\mathbf{exp}(p(1+\varepsilon')\xi)$ and $\mathbf{exp}(p(1+\varepsilon')\xi^{\tau,2})$ follows from Lemmas 13 and 14, respectively. By the particular form of $\Phi(\xi, \xi^{\tau,2})$, it is clear that using Hölder inequality we can show that (6.3) holds, provided that

$$\sup_{r_1, \dots, r_a \in [0, t]} \tilde{\mathbb{E}}[|D_{r_1, \dots, r_a}^a \xi_i|^q] < \infty, \quad 1 \leq a \leq k, 0 \leq i \leq d_Y \tag{6.9}$$

$$\sup_{r_1, \dots, r_a \in [0, t]} \tilde{\mathbb{E}}[|D_{r_1, \dots, r_a}^a \xi_i^{\tau,2}|^q] < \infty, \quad 1 \leq a \leq k, 0 \leq i \leq d_Y \tag{6.10}$$

for any $q \geq 1$. We shall prove the case $i = 1$, the case $i = 0$ being similar, and we will drop the index i in what follows. By Faà di Bruno's formula

$$\begin{aligned} D_{r_1, \dots, r_a}^a \xi &= \int_0^t D_{r_1, \dots, r_a}^a h(X_s) dY_s \\ &= \int_0^t \left(\sum_{l=1}^a h^{(l)}(X_s) B_{a,l}(D_{r_1}^1 X_s, D_{r_1, r_2}^2 X_s, \dots, D_{r_1, \dots, r_a}^a X_s) \right) dY_s. \end{aligned}$$

Hence, by Burkholder-Davis-Gundy inequality, we get

$$\begin{aligned} &\tilde{\mathbb{E}}[|D_{r_1, \dots, r_a}^a \xi|^q] \\ &\leq C(a, q, t) \\ &\quad \times \sum_{l=1}^a \int_0^t \tilde{\mathbb{E}}[|h^{(l)}(X_s) B_{a,l}(D_{r_1}^1 X_s, D_{r_1, r_2}^2 X_s, \dots, D_{r_1, \dots, r_a}^a X_s)|^q] ds \\ &\leq C(a, q, t) \\ &\quad \times \|h\|_{\infty, 2}^q \sum_{l=1}^a \int_0^t \tilde{\mathbb{E}}[|B_{a,l}(D_{r_1}^1 X_s, D_{r_1, r_2}^2 X_s, \dots, D_{r_1, \dots, r_a}^a X_s)|^q] ds, \end{aligned}$$

where

$$\|h\|_{\infty, 2} \triangleq \sum_{l=0}^2 \|h^{(l)}\|_{\infty} < \infty,$$

because $h \in C_b^2$. Therefore, using a generalized version of Hölder inequality and Lemma 15 we get (6.9).

On the other hand, by Leibniz's rule and the Burkholder-Davis-Gundy inequality, we get

$$\begin{aligned} &\tilde{\mathbb{E}}[|D_{r_1, \dots, r_a}^a \xi^{\tau, 2}|^q] \\ &= \tilde{\mathbb{E}} \left[\left| \sum_{\beta \in \mathcal{M}_1(S_0)} \int_0^t D_{r_1, \dots, r_a}^a (L^\beta h(X_{\tau(s)}) I_\beta(\mathbf{1})_{\tau(s), s}) dY_s \right|^q \right] \\ &= \tilde{\mathbb{E}} \left[\left| \int_0^t \sum_{\beta \in \mathcal{M}_1(S_0)} \sum_{l=0}^a \binom{a}{l} (D_{r_1, \dots, r_l}^l L^\beta h(X_{\tau(s)}) (D_{r_1, \dots, r_{a-l}}^{a-l} I_\beta(\mathbf{1})_{\tau(s), s}) dY_s \right|^q \right] \\ &\leq C(m, q, a, t) \\ &\quad \times \int_0^t \sum_{\beta \in \mathcal{M}_1(S_0)} \sum_{l=0}^a \binom{a}{l}^q \tilde{\mathbb{E}}[|(D_{r_1, \dots, r_l}^l L^\beta h(X_{\tau(s)}) (D_{r_1, \dots, r_{a-l}}^{a-l} I_\beta(\mathbf{1})_{\tau(s), s})|^q] ds, \end{aligned}$$

and the proof is further reduced to show for any $\beta \in \mathcal{M}_1(S_0)$ and that

$$\sup_{r_1, \dots, r_l \in [0, t]} \tilde{\mathbb{E}}[|D_{r_1, \dots, r_l}^l L^\beta h(X_{\tau(s)})|^q] < \infty, \quad 0 \leq l \leq a, \quad (6.11)$$

$$\sup_{r_1, \dots, r_l \in [0, t]} \tilde{\mathbb{E}}[|D_{r_1, \dots, r_l}^l I_\beta(\mathbf{1})_{\tau(s), s}|^q] < \infty, \quad 0 \leq l \leq a. \quad (6.12)$$

The proof of (6.11) is similar to the proof of (6.2). The proof of (6.12) is based on the well known fact, see Proposition 1.2.7 and exercise 1.2.5. in Nualart [8], that $D_{r_1, \dots, r_l}^l I_\beta(\mathbf{1})_{\tau(s), s}$ can be expressed as linear combinations of iterated integrals of order lower than l . Then, the result follows from Lemma 5.7.5. in Kloeden and Platen. \square

References

- [1] BAIN, A. and CRISAN, D. (2008). *Fundamentals of Stochastic Filtering*, Stoch. Model. Appl. Probab., Vol 60, Springer Verlag.
- [2] CRISAN, D. (2011). Discretizing the Continuous Time Filtering Problem. Order of Convergence. In *The Oxford Handbook of Nonlinear Filtering*. Oxford Univ. Press, Oxford.
- [3] CRISAN, D. and ROZOVSKY, B. (2011). *The Oxford handbook of nonlinear filtering*, Oxford Univ. Press, Oxford, 2011.
- [4] CRISAN, D. and ORTIZ-LATORRE, S. (2013). A Kusuoka-Lyons-Victoir particle filter. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **469**, no. 2156.
- [5] DEL MORAL, P. (2004). *Feynman-Kac formulae. Genealogical and interacting particle systems with applications*. Probab. Appl. (New York). Springer-Verlag, New York.
- [6] DOUCET, A., DE FREITAS, N. and GORDON, N. (2001). *Sequential Monte Carlo methods in practice*, Springer-Verlag, New York, 2001.
- [7] KLOEDEN, P. and PLATEN, E. (1992). *Numerical solution of stochastic differential equations*. Stoch. Model. Appl. Probab. vol. **23**, Springer-Verlag, Berlin.
- [8] NUALART, D. (2006). *The Malliavin Calculus and Related Topics*. Springer 2006.
- [9] ORTIZ-LATORRE, S. A second order particle filter approximation of the non-linear filtering problem. *Work in progress*.
- [10] PICARD, J. (1984) Approximation of nonlinear filtering problems and order of convergence. Filtering and control of random processes (Paris, 1983), 219–236, *Lecture Notes in Control and Inform. Sci.*, **61**, Springer, Berlin.